


The 23rd Annual International Conference on Auditory Display
Sound in Learning
The Pennsylvania State University, University Park Campus, June 19-23, 2017

Sound in Learning

Proceedings
of the 23rd Annual

International Conference on Auditory Display

The Pennsylvania State University
University Park Campus
June 19-23, 2017

ISBN 0-9670904-4-X

http://icad.org/icad2017/

https://doi.org/10.21785/icad2017.000

Sound in Learning

Proceedings

of the 23rd Annual

International Conference on Auditory Display

The Pennsylvania State University
University Park Campus
June 19-23, 2017

ISBN 0-9670904-4-X

http://icad.org/icad2017/

DOI: 10.21785/icad2017.000

Table of Contents

EXTENDED ABSTRACTS/POSTER PRESENTATIONS

INSTALLATIONS

CONCERT

PRELIMINARIES

ICAD 2017
Organizing Committee

Conference Chairs	Mark Ballora, Jeffrey Rimland
Paper Reviewing Chair	Margaret Schedel
Paper Reviewing Sub-Chairs	Robert Martin (Games session) Neil Verma (Philosophy/Aesthetics session)
Student ThinkTank	Hiroko Terasawa
Workshops	Derek Brock
Installations	Myounghoon Jeong
Music Reviewing Chair	Robert Martin
Technical Director	Curtis Craig
Workshop on Issues of Diversity, Equity, and Inclusion	Areti Andreopoulou
Proceedings Editor	Margaret Schedel, Mark Ballora
Website	Mark Ballora
Moral Compass	Matti Gröhn
Liaisons, Penn State Conferences & Institutes	Anu Mathew, Cheryl Rae Stamm

ICAD Board (2017)

Executive Members

David Worrall, *president* (Columbia College Chicago, USA)
Derek Brock, *treasurer* (Naval Research Laboratory, USA)
Areti Andreopoulou, *secretary* (LabMAT, University of Athens, Greece)

Emeritus Members

Gregory Kramer, *founder* (Clarity and Metta Foundation, USA)
Matti Gröhn (Professional Cooperative Polkuvekosto, Finland)

Elected Members

Brian Katz (LIMSI-CNRS, France)
Paul Vickers (Northumbria University, United Kingdom)
Marc St. Pierre *student representative* (Simon Fraser University, Canada)

Ex-Officio Members

Mark Ballora (The Pennsylvania State University, USA)
Visda Goudarzi (Institute of Electronic Music and Acoustics, Austria)
S. Camille Peres (Texas A&M University, USA)
Hiroko Terasawa (University of Tsukuba, Japan)

Welcome to ICAD from the Conference Chairs

Mark Ballora and Jeffrey Rimland

The year 2017 marks a quarter century from the first ICAD. The 23rd International Conference of Auditory Display offers an occasion to reflect on the progress that has been made since that first gathering in 1992 in Santa Fe. As a leading research university and a strong proponent of interdisciplinary research, Penn State is proud to be hosting this year's conference. Thanks to the sponsorships of the Office of the Vice President for Research and the College of Arts and Architecture's ADRI (Arts and Design Research Incubator), a broad cross-section of Penn State researchers will be in attendance. Their insights will undoubtedly enrich the conference proceedings as well as lay the groundwork for further work in the area to be carried out at Penn State. Additional support from the Department of Acoustics and the College of Information Sciences and Technology demonstrates Penn State's commitment to furthering the field of auditory display and sonification.

This year's theme is Sound in Learning, with an emphasis on the role of sonification in science pedagogy. If a generation of students is raised to regard science as something that is heard as well as seen, will they possess a more holistic and intuitive understanding of material than they would get from visual materials alone?

To explore this question, the conference features a healthy blend of arts and science, theoretical and applied, technical and creative. Events include traditional paper and poster sessions, plus a diverse range of workshops, installations, and concert pieces. Special events will also include a sonification contest in EEG data sponsored by the University of Waterloo and Goldsmiths, University of London, a workshop in issues of diversity, equity and inclusion, and a Do-A-Thon, which will explore ways to increase the visibility and viability of sonification on the web. As in years past, the National Science Foundation is sponsoring a graduate student ThinkTank, a pre-conference symposium that allows emerging students to share their work and receive mentoring from senior members of the field. The keynote speakers will include composer and software designer Carla Scaletti, an alumna of the first ICAD and the person who gave us the first definition of the term "sonification," and Elizabeth Cohen, a leading sound and acoustical consultant in the entertainment industry.

The setting for Penn State is an area nicknamed "Happy Valley" in the 1930s, when the university's presence insulated the area from some of the harsher circumstances of the Great Depression. State College is a classic "college town," featuring a variety of shopping, restaurants, musical venues, and recreational opportunities. From its founding as the Farmer's High School in 1855, to its establishment as a land grant university with an agricultural mission in 1863, to its emergence as a major center for applied research in the 1940s, Penn State has evolved into one of the world's leading research universities. Situated within the beauty of the central Pennsylvania foothills, the university has long set standards of excellence in higher education, and has one of the largest and most loyal alumni associations in the world.

We hope that you enjoy your time here. And we hope that afterwards you will understand why Penn State alumni greet each other worldwide with the proud affirmation: "We are...Penn State!"

Conference Theme
Sound in Learning

Scientific literacy is typically gained through the study of graphs and various types of visualizations. Many of these have been in existence since the late 18th century, and are part of the standard research vocabulary.

The twentieth and twenty-first century have made dynamic, multi-modal displays feasible. Visualization is essential for many applications — it draws on the strength of the eyes for assessing static qualities such as size, color, or texture. But many applications could greatly benefit from displays the address the ears, with their particular sensitivity to dynamic changes and capability for following multiple simultaneous streams.

Auditory information is also received faster than visual information. Hearing sets the stage for what we see. Sound is quickly transmitted to areas of the brain that carry out basic functions at an emotional, survival level. The legacy of our ancestors' quick "fight or flight" response is the human creature's unique appreciation of music.

Because of all this, sound should be a part of learning science and other topics. Young students being introduced to information through sound will likely have a more holistic and engaging experience than is possible with visual materials alone. If a generation of students were raised to learn about science by listening as well as looking, what implications would this have for the scientific climate twenty or thirty years in the future?

ICAD 2017 Program

Monday, June 19

9:00 - 5:00 Student ThinkTank – 201 IST Building

Tuesday, June 20

9:00 - 1:00 Workshops	
Takahiko Tsuchiya	Live Coding Sonification System for Web Browsers with Data-to-Music API **Location: 210 IST Building**
Matthew Neal and Nicholas Ortega	Tutorial on Higher Order Ambisonics and demonstration of the Auralization and Reproduction of Acoustic Sound-fields (AURAS) facility at Penn State **Location: 214 Applied Science Building**
Myounghoon Jeon, S. Maryam FakhrHosseini, Eric Vasey	New Opportunities for Auditory Interactions in Highly Automated Vehicles **Location: 201 IST Building**
Installations open: 9:00-5:00 125 Borland, 10:00-2:00 16 Borland	
1:00 Registration – IST Building Cybertorium Lobby	
2:00 Opening Welcome - IST Building Cybertorium	
3:00 - 4:35: Paper Session 1 – Language - IST Building Cybertorium	
Thomas Gable, Brianna Tomlinson, Stanley Cantrell and Bruce Walker	Spindex and Spearcons in Mandarin: Auditory Menu Enhancements Successful in a Tonal Language
Michael Nees, Joanna Harris and Peri Leong	How do people think they remember melodies and timbres? Phenomenological reports of memory for nonverbal sounds
Areti Andreopoulou and Visda Goudarzi	Reflections on the Representation of Women in the International Conferences on Auditory Displays (ICAD)
Daniel Verona and Camille Peres	A Comparison Between the Efficacy of Task-Based vs. Data-Based sEMG Sonification Designs
Stephen Taylor	From Program Music to Sonification: Representation and the Evolution of Music and Language
5:00 - 6:00 Opening Reception - IST Building Cafe	
6:00 - 7:00 Keynote 1: Carla Scaletti - IST Building Cybertorium	

Wednesday, June 21

8:30 – 9:00: Registration/Continental Breakfast – IST Building Cybertorium Lobby

Installations open: 9:00-5:00 125 Borland, 1:00-5:00 16 Borland

9:00 - 10:20: Paper Session 2 – Movement - IST Building Cybertorium

Joseph Newbold, Nicolas Gold and Nadia Bianchi-Berthouze	Musical Expectancy in Squat Sonification for People who Struggle with Physical Activity
Jon Bellona, Luke Dahl, Amy LaViers, Lin Bai	Empirically Informed Sound Synthesis Application for Enhancing the Perception of Expressive Robotic Movement
Jason Sterkenburg, Steven Landry and Myounghoon Jeon	Influences of Visual and Auditory Displays on Aimed Movements Using Air Gesture Controls
Juliana Cherston and Joseph A. Paradiso	Rotator: Flexible Distribution of Data Across Sensory Channels

11:00 - 12:00 Keynote 2: Elizabeth Cohen- IST Building Cybertorium

12:00 – 1:00: Lunch – IST Building Cybertorium Lobby

1:00 - 2:20 Paper Session 3 - Navigation & Noise - IST Building Cybertorium

Joseph Schlesinger, Brittany Sweyer, Alyna Pradhan and Elizabeth Reynolds	Frequency-Selective Silencing Device for Digital Filtering of Audible Medical Alarm Sounds to Enhance ICU Patient Recovery
Ruta Sardesai, Thomas Gable and Bruce Walker	Introducing Multimodal Sliding Index: Qualitative Feedback and Perceived Workload for Auditory Enhanced Menu Navigation Method
Woodbury Shortridge, Thomas Gable, Brittany Noah and Bruce Walker	Auditory and Head-Up Displays for Eco-Driving Interfaces
Kees van den Doel and Michael Robinson	Use of sonification of RADAR data for noise control

Afternoon – Light snacks available in the Borland Building

2:30 - 3:00 Open session with ICAD Board
3:00 - 5:30 ICAD Board Meeting
Room 121 Borland Building

2:30 - 5:30 Sonification Do-a-thon - Room 113 Borland Building

5:30 - 7:00 Poster Session - Playhouse Theatre Lobby – Light snacks available

Marlene Mathew	BSONIQ: A 3-D EEG Sound Installation
Yuanjing Sun, Jaclyn Barnes and Myounghoon Jeon	Multisensory Cue Congruency in the Lane Change Test
Marcelo Ferranti and Rejane Spitz	Sounding objects: an overview towards sound methods and techniques to explore sound within design processes
Peter Coppin, David Steinman, Daniel MacDonald and Richard Windeyer	Progress Toward Sonifying Napoleon's March and Fluid Flow Simulations through Binaural Horizons
Paulo Marins	Challenges and Constraints of Using Audio in Online Music Education
Wenyu Wu, Alexander Gotsis, Rudina Morina, Harsha Chivukula, Arley Schenker, Madeline Gardner, Felix Liu, Spencer Barton, Steven Woyach, Bruno Sinopoli, Pulkit Grover and Laurie Heller	Echoexplorer: A Game App for Understanding Echolocation and Learning to Navigate Using Echo Cues

Wednesday, June 21, cont.

7:00 - 8:00 Concert - Playhouse Theatre	
Stephen Roddy	Sonification: The Good Ship Hibernia (audio)
Alfredo Ardia	Rami (video)
Julius Bucsis	Portraits of Nine Revolving Celestial Spheres (audio)
Roberto Zanata	After Images (video)
Antonio D'Amato	Körper (multichannel audio)
James Cave and Ben Eyes	Eonsounds: Fiamignano Gorge (voice and tape)
Andrew Litts	Singularity for trumpet and electronics

Thursday, June 22

8:30 – 9:00: Registration/Continental Breakfast – IST Building Cybertorium Lobby	
Installations open: 9:00-5:00 in 125 Borland, 10:00-2:00 in 16 Borland	
9:00 - 10:20 - Paper Session 4: Ecology - IST Building Cybertorium	
Josh Laughner, Elliot Canfield-Dafilou	Illustrating trends in nitrogen oxides across the United States using sonification
Brianna Tomlinson, R. Michael Winters, Chris Latina, Smruthi Bhat, Milap Rane and Bruce Walker	Solar System Sonification: Exploring Earth and Its Neighbors through Sound
Kelly Fox, Jeremy Stewart and Rob Hamilton	MADBPM: Musical and Auditory Display for Biological Predictive Modeling
Arthur Pate, Benjamin Holtzman, John Paisley, Felix Waldhauser and Douglas Repetto	Pattern Analysis in Seismic Data Using Human and Machine Listening
11:00 - 12:00 - Paper Session 5: Games - IST Building Cybertorium	
James Broderick, Jim Duggan and Sam Redfern	Using Auditory Display Techniques to Enhance Decision Making and Perceive Changing Environmental Data within a 3D Virtual Game Environment
Adrian Jäger and Aristotelis Hadjakos	Navigation in an audio-only first person adventure game
Laurie Heller, Arley Schenker, Pulkit Grover, Madeline Gardner and Felix Liu	Evaluating two ways to train sensitivity to echoes to improve echolocation
12:00 – 1:30: Lunch – IST Building Cybertorium Lobby	
1:30 - 2:50 - Paper Session 6: Philosophy/Aesthetics - IST Building Cybertorium	
Takahiko Tsuchiya and Jason Freeman	Spectral Parameter Encoding: Towards a Framework for Functional-Aesthetic Sonification
Parisa Alirezaee, Roger Girgis, Taeyong Kim, Joseph Schlesinger and Jeremy Cooperstock	Did you feel that? Developing Novel Multimodal Alarms for High Consequence Clinical Environments
Steven Landry and Myounghoon Jeon	Participatory Design Research Methodologies: A Case Study in Dancer Sonification
3:00 - 4:30 – Open Forum on Issues of Diversity, Equity and Inclusion - 201 IST Building	
6:00 - 8:00 – Banquet - IST Building Cafe	

Friday June 23

9:00 – 9:30: Continental Breakfast – IST Building Cybertorium Lobby	
Installations open: 9:00-2:00 125 Borland, 10:00-2:00 16 Borland	
9:30 - 10:45 - Paper Session 7: Computing - IST Building Cybertorium	
David Worrall	Computational Designing for Auditory Displays
Jiajun Yang and Thomas Hermann	Parallel Computing of Particle Trajectory Sonification to Enable Real-Time Interactivity
Brian Bartling and Catherine Psarakis	Simulation Products and the Multi-Sensory Interactive Periodic Table
Samuel Chabot and Jonas Braasch	An Immersive Virtual Environment for Congruent Audio-Visual Spatialized Data Sonifications
11:00 - 12:00 - Open Mic/Closing - IST Building Cybertorium	

Keynote 1
Carla Scaletti - Tuesday, June 20, 6:00 PM

Software developer / composer Carla Scaletti gave an invited talk at the very first ICAD, organized by Greg Kramer in 1992, based on work she and Alan Craig had carried out the previous year at the National Center for Supercomputing Applications, creating data-sonifications to accompany data-visualization videos produced at NCSA ("Using Sound to Extract Meaning from Complex Data" SPIE 1991). The subsequent year, Scaletti, Brian Evans and Robin Bargar presented a pre-conference tutorial on data sonification for SIGGRAPH-93.

More recently, she's collaborated with CERN physicist Lily Asquith on the LHCSound project to sonify data from the Atlas Experiment (2010 - 2012) and with Swiss choreographer Gilles Jobin on a soundtrack for his physics-inspired dance piece QUANTUM incorporating data-driven sound (2012). Last year, Scaletti was invited to contribute a chapter for the forthcoming *Oxford Handbook of Algorithmic Composition* somewhat controversially entitled "Sonification ≠ Music."

Scaletti's primary focus is the creation of software tools for real-time sound synthesis and manipulation in the Kyma sound design language at Symbolic Sound Corporation (which she co-founded with Kurt J. Hebel), while continuing to create both data-driven music and scientific data sonifications (which she insists are two very different activities). Along the way, she's discovered that doing data sonification changed the way she thinks about data, and about mapping, and about sound synthesis and, somewhat unexpectedly, it's also changed the way she thinks about music.

She has a doctorate in music and a master's of computer science from the University of Illinois where she studied composition with Salvatore Martirano, John Melby, Herbert Brün and Scott Wyatt and computer science with Ralph Johnson, one of the Design Patterns "Gang of Four."

In 2015, she was invited to present a keynote address at the International Computer Music Conference (ICMC 2015) and was an invited lecturer at GVA Sessions — a workshop in Geneva involving choreographers, filmmakers, and particle physicists from CERN. Each year, she co-organizes the Kyma International Sound Symposium (KISS) which, in 2017 will be in Oslo Norway on the theme "Augmenting Reality".

For more information, go to carlascaletti.com.

Keynote 2
Elizabeth Cohen - Wednesday, June 21, 11:00 AM

Elizabeth Cohen, engineer for the arts, received her M.S. degree in Electrical Engineering and Ph.D. in Acoustics from Stanford University. She has served on the faculty of Stanford University since 1980 as Consulting Professor of Electrical Engineering at Stanford and as a visiting researcher at the Center for Computer Research in Music and Acoustics (CCRMA). She was also appointed as a Visiting Professor of Information Studies and Theater, Film, and Television at UCLA and a member of the Moving Image Archive Studies Program.

In 1982, she founded Cohen Acoustical Inc. Her clients have included The Academy of Motion Picture Arts and Sciences, CBS, Digital Theater Systems, Dolby Labs, Fraunhofer Institute, Grateful Dead Productions, The Los Angeles Philharmonic, Meyer Sound, The Walt Disney Company, Sony, Universal Studios and numerous other arts and entertainment organizations. Her acoustical design of the Joan and Irving Harris Concert Hall in Aspen, Colorado has received worldwide acclaim.

From 1993-4, Dr. Cohen served as Science and Engineering Fellow to the White House Economic Council where she was responsible for Arts and Information Infrastructure Initiatives. In 1998, she received the Touchstone Award, from Women in Music and the Distinguished Alumni Award from the Stanford Women's Center. In 2000 she received the AES Distinction Award, for Pioneering the Technology Enabling Collaborative Performance over the Broadband Internet. As part of the CineGRid and broadband network interdisciplinary community her current work is focused on the research, development, and demonstration of networked collaborative tools to enable the production, use and exchange of very-high-quality digital media over photonic networks.

She is a Past President and Fellow of the Audio Engineering Society, Member of the Academy of Motion Picture Arts and Sciences (AMPAS), Fellow of the Acoustical Society of America, member of the Society of Motion Picture and Television Engineers, Ad Hoc Postproduction Committee. She has served on the National Academy of Recording Arts and Sciences Committees on Archiving and Preservation and the Producers and Engineers Wing Advisory Council. She is currently serving on the Sound Preservation Board of the Library of Congress, Rhythm for Life Foundation, and the Institute for Music and Neurologic Function. In addition, she is the Vice Chair of Education for the AMPAS Science and Technology Council, a member of its Digital Motion Picture Archive Committee, and is the co-chair of The Academy Archive Digital Content project. Dr. Cohen served on the National Science Foundation Blue Ribbon Panel on Sustainable Digital Preservation and Access and the Sloan Foundation Stewardship Gap Advisory Group.

Dr. Cohen has collaborated with Nobel Laureate, George Smoot, on outreach efforts to expand student understanding of the role of science and technology in the creative arts. She continues her affiliation with numerous STEAM programs to build inclusive pipelines for the next generation of creative engineers.

WORKSHOPS

Takahiko Tsuchiya
Live Coding Sonification System for Web Browsers with Data-to-Music API

Today's web environment allows us to build highly accessible web applications for data sonification with powerful real-time audio synthesis. Moreover, in the form of "live coding", the development of a sonification becomes iterative, responsive, and exploratory. In this workshop, the participants will experience web-based live coding of sonification with JavaScript and the Data-to-Music API. First, we cover the basic concepts of web development such as HTML, JavaScript, and file serving (10-20 minutes). We then learn the basic operation of data handling and audio synthesis in the DTM API (20-30 minutes), followed by more advanced techniques for transformation, analytics, and mapping for expressive sonification (20-30 minutes). Then, the participant will have an opportunity to experiment with these techniques using a data set of their choice (30 minutes - 1 hour). We conclude the session with presentations from the participants, and discuss a few more applications of DTM such as the communication from the browser to Max/MSP via web sockets or to a DAW software via Web MIDI API (10-20 minutes).

Matthew Neal and Nicholas Ortega, Adviser: Michelle C. Vigeant
Tutorial on Higher Order Ambisonics and demonstration of the Auralization and Reproduction of Acoustic Sound-fields (AURAS) facility at Penn State

Ambisonics, originally proposed by Gerzon in the 1970's, is a technique to reproduce a measured sound field with an array of loudspeakers. Since then, computer and audio technology has seen vast advances, and it is now possible for researchers and individuals to recreate measured sound fields using Ambisonics and Higher Order Ambisonics (HOA). The spherical harmonic components of a sound field can be measured with a spherical microphone array and reproduced using HOA. Commercially available spherical microphone arrays along with the increasing availability of cost-effective multi-channel audio systems make implementing a HOA system quite accessible. This workshop will provide key background information on the acoustic fundamentals of spherical harmonics, spherical microphone arrays and Ambisonics. With the foundations laid, attendees will learn about the Auralization and Reproduction of Acoustic Sound-fields (AURAS) facility at Penn State. The AURAS facility is a 30-loudspeaker HOA array located within an anechoic chamber on Penn State's campus. Details of the facility's construction, processing techniques, hardware setup, and software implementations will be presented. Current and future research projects utilizing this facility will also be outlined. After the presentation of the AURAS facility, a demonstration of open-source tools for implementing HOA in MATLAB and Max7 will be provided. Attendees will gain a baseline understanding of how to use these tools in implementing a HOA system. Live demonstrations of the AURAS facility will be included at the end of the workshop for attendees.

Both Matthew Neal and Nicholas Ortega are Ph.D. Candidates in the Sound Perception and Room Acoustics Laboratory (SPRAL) out of Penn State's Graduate Program in Acoustics. Working with Dr. Michelle C. Vigeant, they use spherical array processing techniques with both microphone and loudspeaker arrays in subjective and psychoacoustic testing.

Myounghoon Jeon, S. Maryam Fakhr Hosseini, Eric Vasey

New Opportunities for Auditory Interactions in Highly Automated Vehicles

Vehicle automation is becoming more widespread. As automation increases, new opportunities and challenges have also emerged. To name a few, maintaining driver situation awareness, timely cueing take-over-requests, and providing additional cues to pedestrians are all new areas of research. In this workshop we aim to address new opportunities and directions of auditory interactions in highly automated vehicles to provide better driver user experience and to secure road safety.

We have five explicit goals in our workshop:

1) Provide an organized thinking about the topic of how auditory interactions can be efficiently and effectively applied to highly automated vehicle contexts;
2) Build and nurture a new community that bridges auditory display community with automotive user interface community;
3) Discuss and exchange ideas within and across sub-communities;
4) Suggest promising directions for future transdisciplinary work; and
5) Yield both immediate and long-term community-, research-, and design-guidance products.

To this end, we will invite researchers and practitioners from all backgrounds, who are interested in auditory displays and automotive user interface fields. Achieving these goals will provide an opportunity to move this integrated field forward and build a solid community that includes ICAD.

SPECIAL EVENTS

Student ThinkTank

The ICAD Student ThinkTank is a pre-conference symposium for graduate students, funded by special arrangement with the National Science Foundation.

USA-based graduate students doing work in auditory display are invited to submit applications to the ThinkTank. Selected participants present their work in progress to the other participants as well as to a panel of senior researchers in the field. Discussions and breakout sessions allow students to explore productive paths of research and funding, as well as to foster friendships and networking.

ThinkTank Chair

Hiroko Terasawa, Tsukuba University, Japan

ThinkTank students

Andrew Litts, Temple University
Matthew Neal, The Pennsylvania State University
Joseph Newbold, University College London, UK
Kyle Shaw, University of Illinois at Urbana-Champaign
Brianna Tomlinson, Georgia Institute of Technology
Takahiko Tsuchiya, Georgia Institute of Technology
Judy Twedt, University of Washington
Eric Vasey, Michigan Technological University
Daniel Verona, Texas A&M University
Mike Winters, Georgia Institute of Technology

Panelists

Matti Gröhn, Professional Cooperative Polkuvekosto, Finland
Sungyoung Kim, Rochester Institute of Technology
Bruno Ruviaro, Santa Clara University

Open Forum on Issues of Diversity, Equity, and Inclusion

Areti Andreopoulou, chair

In academia, the benefits of "unbiased research evaluation"—that is, the methods by which scholarly works are evaluated in a manner that is objective and free of prejudice—are well understood and respected. Review processes for research proposals and scientific article submissions have defined guidelines that have been widely adopted in an attempt to secure a "fair" process of review.

But the level of true inclusivity in academia is frequently called into question, particularly outside of the formal peer review process, where people's work, ideas, perspectives, and beliefs are not "protected" by anonymity. It is self-evident that people thrive best in environments where they feel free to act and express themselves. But academia often struggles with the issue of how to offer truly diverse environments that are conducive to the personal and professional development of a broad range of populations.

There has never been a more relevant time for academic institutions and organizations to take action towards strengthening diversity, equity, and inclusion. The 23rd International Conference of Auditory Display (ICAD 2017) will offer an open forum where issues pertinent to those topics will be discussed and plans for improvement will be proposed.

ICAD Do-A-Thon

For people not already involved with sonifications in some way, there are few opportunities to engage with the topic. With this ICAD Do-A-Thon, we want to explore ways to increase the visibility and viability of sonification on the web, e.g. by:

- making sonifications available under open licenses and in open formats, so that they can be reused in non-specialist contexts like Wikipedia pages;
- creating sonifications specifically for Wikipedia and its sister projects;
- making the source code for sonifications available in programming environments that can be readily explored by people outside the field, e.g. containerized Jupyter notebooks on the basis of open data.

We invite our community to participate in this activity. The kick-off event will be June 21st at ICAD 2017 at Penn State, but we will make it easy for anyone around the world to engage, and hope that researchers will continue to share their data, code, and sonifications more broadly based on the models we develop.

Sonification Competition: Turning Brainwaves into Sounds

Mike Schaekermann, Edith Law
University of Waterloo (Canada)

Rebecca Fiebrink
Goldsmiths, University of London (UK)

The ICAD17 sonification competition is unique in many ways, and offers exciting opportunities to contribute to applied research. Its theme is "Turning Brainwaves into Sounds." The objective of the competition is to sonify sleep EEG (human brainwave signals captured during sleep), with the goal of creating mappings that help to distinguish signals associated with different sleep stages.

Background

The analysis of sleep EEG is essential for understanding sleep patterns and pathologies of patients. In current practice, medical technicians manually classify thousands of epochs of EEG data into 5 stages of sleep (Awake, N1, N2, N3, REM), a process that is extremely time-consuming and tedious. At University of Waterloo, we are developing new human-in-the-loop intelligent systems to more efficiently classify medical time series data. In a project called CrowdEEG (http://crowdeeg.ca), we study how to leverage crowdsourcing to tackle this problem.

Not surprisingly, it can be challenging to teach non-experts to do the expert-level analysis task typically handled by medical technicians. Our question, and the motivation behind this competition, is:

"By sonifying sleep EEG data, can we enable non-experts to classify the signal into sleep stages, or identify transitional points between sleep stages, simply by listening?"

Sonifying EEG is a non-trivial process, requiring expertise in sound engineering --- which is where you and the rest of the ICAD community come into play!

Objective

The goal is to maximize the within-class similarities (e.g., make all the REM sleep epochs sound similar) and between-class distances (e.g., make the REM sleep and Awake epochs sound different), so that the crowd of non-experts can easily distinguish between the 5 classes, and identify transitional points between sleep stages.

Procedure

Participants will be provided with (1) a web-based sonification tool that allows them to construct a mapping from time series to sound, (2) a dataset containing sleep EEG data, consisting of several types of biosignal time series, including brainwave activity.

Evaluation

Instead of relying on expert jury members, we will use a crowdsourced approach to determine the best mapping. Each submitted sonification will be tested on a large number of novices via the crowdsourcing platform Mechanical Turk, and the mapping with the highest average classification performance among novices will win the competition.

PAPER PRESENTATIONS

Paper Session 1
Language

SPINDEX AND SPEARCONS IN MANDARIN: AUDITORY MENU ENHANCEMENTS SUCCESSFUL IN A TONAL LANGUAGE

Thomas M. Gable

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
thomas.gable@gatech.edu

Brianna Tomlinson

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
btomlin@gatech.edu

Stanley Cantrell

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
cantrell@gatech.edu

Bruce N. Walker

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
bruce.walker@psych.gatech.edu

1. ABSTRACT

Auditory displays have been used extensively to enhance visual menus across diverse settings for various reasons. While standard auditory displays can be effective and help users across these settings, standard auditory displays often consist of text to speech cues, which can be time intensive to use. Advanced auditory cues including spindex and spearcon cues have been developed to help address this slow feedback issue. While these cues are most often used in English, they have also been applied to other languages, but research on using them in tonal languages, which may affect the ability to use them, is lacking. The current research investigated the use of spindex and spearcon cues in Mandarin, to determine their effectiveness in a tonal language. The results suggest that the cues can be effectively applied and used in a tonal language by untrained novices. This opens the door to future use of the cues in languages that reach a large portion of the world's population.

2. INTRODUCTION

The modern computer interface primarily relies on the use of the visual modality to relay information to the user. However, the employment of visuals is not always viable for users, whether it is due to vision impairment, lack of visual clarity from a system because of increasingly smaller screens, or situational blindness such as when the user is driving or completing another visually-demanding task. In these instances the transfer of information to the user via auditory displays can often be a viable option. Examples of auditory displays range from fire alarms and other basic interfaces to complex computer systems. These systems can employ speech-based or non-speech audio to relay information to users. Extensive research has shown the successful deployment of auditory displays for user interfaces ranging from warning systems to menus.

In the current document we explore the ability of novices to use two popular advanced auditory cues (spindex and spearcons) in Mandarin, a tonal language in which these advanced cues have not yet been extensively evaluated.

3. AUDITORY MENUS

Menus are a part of many interfaces that we interact with on a daily basis. Their complex structure and the difficulty in navigating through them while understanding what they contain can pose a large problem in the employment of auditory interfaces. Individuals with vision impairment rely on auditory menus and displays to interact with user interfaces for computers, phones, and other technologies; these devices rely on screen readers, which use text-to-speech (TTS) output to provide contextual description for the software content. The TTS interfaces employed in these screen readers are one of the simplest forms of auditory feedback from a computer interface and have been used for many types of auditory displays to support accessibility. In addition, auditory menus have become widely used in hands-free calling or other contexts when a user may be situationally visually impaired, such as when driving or doing other complex visual-manual tasks [1].

3.1. Speech-based interfaces

These speech-based interfaces have been shown to relay information to the users fairly well. However, TTS and other interfaces are limited in their abilities to provide a fast interaction and cannot always relay all details to the users successfully. Previous work has shown that TTS interfaces can lead to slower interaction times than visual interaction [2] [3]. One study found that in a dual-task situation when users were asked to complete a search task while driving they employed the TTS auditory menus until the importance of the secondary task was heightened, at which point the users abandoned the auditory display and relied on visuals to

https://doi.org/10.21785/icad2017.025

complete the task [4]. These findings show that when users are given a secondary task rated with a higher importance than other tasks (and have the ability to choose interface modality), they will choose to use their visual attention to complete the task if the interaction time is too long using an auditory displays, even if it leads to more dangerous primary task behaviors.

3.2. Non-speech interfaces

In order to address the issue of slow feedback in TTS interfaces, researchers have developed a number of non-speech auditory approaches to transfer information to the user, including auditory icons, earcons, and audemes. Auditory icons are sounds which portray real world concepts and objects directly and usually have a one to one mapping, such as the sound of a *creaking door* representing *entering a new room* or *opening a file* [5]. Auditory icons leverage previous knowledge on the part of the listener, and provide an easy way to transfer knowledge to the designed sound. Earcons are another type of non-speech auditory display, but instead of mimicking or directly representing real-world sounds, they are purposefully designed, musical tones [6]. Audemes present a more complex representation through layers of multiple sets of music or sound effects to portray more complex themes or concepts to support educational contexts [7]. Though these types of displays can be easily recognizable and learnable, they can potentially lead to large amount of memorization for users, and are not as usable for supporting menu or long lists of options.

3.3. Advanced auditory cues

To support faster navigation of auditory menus compared to typical TTS menus, several types of advanced auditory cues have been created, including spearcons and spindex. *Spearcons* are unique, algorithmically-condensed pieces of speech, which are based on their original TTS words or phrases, but are often no longer recognizable as a specific spoken utterance [8] [9]. Previous work has found that spearcons reduce learning time for an auditory menu compared to other non-text auditory feedback such as earcons [10]. *Spindex* cues are a set of sounds that represent the initial sounds of menu items, to support faster searching through a large auditory list or menu (e.g., a listener might use the spindex to skip past As, Bs, Cs, and so on, until they reach T, which they know is the first letter of a song title) [11]. Spindex cues have been shown to decrease search time compared to plain TTS; and users report them to have a high level of helpfulness compared to other speech-based auditory menus [11][12].

Though spindex and spearcons were both initially developed in English, some researchers have explored using spearcons and spindex to enhance TTS interfaces across multiple languages including German, Hungarian, Korean, and Swedish [13][12][14]. Recently, researchers have evaluated how well Mandarin spearcons support eyes-free navigation of real physical environments (i.e., not used in a user interface) compared to English spearcons and to plain Mandarin TTS (with no enhancements) [15]. They found that Mandarin spearcons are better at conveying distance from a target, but the methodology of that study relied heavily on training for both types of spearcons to scaffold initial learning, and had a limited set of spearcons that the listeners used to navigate. Something which was not shown was whether or not spearcons are successful for user interfaces in tonal languages (such as Mandarin) without this extensive training and with a larger set of cues. Spearcon-shortening of speech sounds might affect the tonal pronunciation for a broad range of words, like those which might show up in a long playlist for songs. In addition, there has not been extensive exploration of spindex cues in languages other than English, particularly in tonal languages. We explore both of these concepts through our two experiments below.

4. CURRENT STUDY

The current studies investigated the ability of spindex and spearcon cues to work in a tonal language for long list and multiple tab menu navigation with no training. Two separate studies were undertaken: one for menu navigation of long lists, and another for navigation of multiple-tab menus.

It was expected that both auditory cues would work in Mandarin, be recognizable by participants, and assist in navigation without any training necessary as seen through either a lack of any differences in performance between the TTS and advanced auditory cues, or seeing better performance with the advanced auditory cues.

5. STUDY 1

5.1. Participants

A total of 23 participants (15 males and 8 females) with an average age of 22.1 (SD = 4.5) from a large research university in the United States took part in the study. Only native speakers of standard Mandarin were recruited for the study, and all participants were required to have normal or corrected to normal vision and hearing. Participants reported having spoken Mandarin for an average of 20.4 years (SD=5.9) and writing the language for an average of 18.2 years (SD=5.5). The mean self-reported fluency ratings for Mandarin on a scale from 1-6 (6 being the highest fluency) was 5.9 (SD=0.3), and for writing it was 5.7 (SD=0.9). All participants signed informed consent and provided demographic information; they also completed a few questions regarding their knowledge of Mandarin to ensure an equivalent minimum level of Mandarin expertise.

5.2. Apparatus

Visual stimuli were presented using a 21.5 inch monitor with 1440 x 900 pixel resolution; auditory stimuli were presented using Sony MDR-7506 Studio headphones. Participant responses were collected in a sound attenuated chamber to isolate the sounds used in the research, and to ensure no other noises competed for the participants' attention. A software program written in JAVA and using the APWidgets library [16] was run from an iMac computer, with a 2GHz Intel Core 2 Duo processor and 1GB of RAM running Mac OSX 10.10.4 and displayed on the 21.5 inch monitor described above. The software was created to randomize across a block system, cue participants to when a task needed to be completed, collect responses, and record data. Participants used the connected keyboard to input their responses to the computer, which was placed on a desk in front of them.

To measure subjective workload the common measure of NASA Task Load Index (NASA-TLX) was used. The index measures six subscales of effort including effort, temporal

demand, physical demand, frustration, performance, and mental demand [17].

5.3. Menu structures

The song list menu in this study was created by taking SBS POP ASIA 2015 top 100 songs in China and removing titles that were in English, or translating the English song titles to Mandarin, as appropriate. This left 94 songs in total, which were then sorted alphabetically using Mandarin pinyin ordering. The menu was navigated using the arrow keys on the keyboard, selecting a song with the enter key.

5.4. Auditory stimuli

The auditory stimuli were created by recording a female native Mandarin speaker reading out all of the menu items used in the study. This was done due to the text-to-speech (TTS) generators currently available on the market being rated poorly and being said to sound "unnatural" by native speakers in an initial pilot. These human recordings were then put through a number of algorithms to create the necessary spearcon and spindex cues for the menu system. The spearcon cues were created by a C++ program used to create a linear logarithmic compression of the TTS audio (.WAV) files. Spindex cues for the study were determined by the pinyin of each song item and each unique pinyin was recorded as individual audio (.WAV) files from the female native Mandarin speaker. These audio files were then stitched together using Sound eXchange to create the more complex auditory cues [18].

5.5. Procedure

Upon arrival participants were asked to confirm they met the study criteria and then read through and signed an informed consent form. Then, the participants were assigned to the order of which menu structure they completed first, which was randomized between participants.

Participants were then introduced to the structure of the menu by looking at and interacting with it visually. No audio was provided as they did this, and participants used the arrow keys to move up and down the list, as practice for the study block. Participants were also able to see where the target item would be shown during the experiment. Next, the participants started the condition blocks; these consisted of a training phase of 5 random item selections followed by 20 items for each condition. These 20 items were selected via a semi-random bin system where the total list was divided into 4 bins and one song was pulled from each bin before repeating a bin. This was done to ensure choices throughout the entire list instead of the possibility of an uneven distribution. There were 5 conditions in the study including Visual-only, TTS, Spindex+TTS, Spearcon+TTS, and Spindex+Spearcon+TTS. After each condition participants completed the NASA-TLX to measure subjective workload and moved onto the next condition in the block. In total there were 3 blocks of each of the 5 conditions, resulting in 15 total blocks.

After completing all 15 blocks of menu search tasks, participants completed a demographics survey about preferences for the cues, perceived fun of use, likability, annoyance, helpfulness, and effectiveness of the cues.

5.6. Analysis

The data for time to complete each selection task, accuracy, and TLX scores were collected for each block. All of these data were analyzed with within-subject repeated measures ANOVA, with Huynh-Feldt corrections, as appropriate, for violations of sphericity assumptions. The survey data were analyzed via basic paired t-tests.

5.7. Results

5.7.1. Interaction (Search) Time

The data for interaction (search) time are shown in Figure 1 (broken into conditions to allow for more detail of the conditions). A two way repeated measures ANOVA with Huynh-Feldt corrections was done on the interaction time results, and revealed a significant main effect of block, $F(1.67, 35.05) = 109.98$, $p < .001$. To determine the differences in interaction time between blocks, post-hoc comparisons of 3 t-tests with Bonferroni corrections (correcting alpha to .0166) were performed. The post-hocs revealed significant differences between Blocks 1 and 2, $t(22) = 10.41$, $p < .001$; Blocks 1 and 3, $t(21) = 12.42$, $p < .001$; and Blocks 2 and 3, $t(21) = 4.39$, $p < .001$.

The original two-way repeated measures ANOVA found no significant main effect of condition $F(3.98, 83.51) = 0.96$, $p = .435$, and no significant interaction, $F(5.55, 116.60) = 1.49$, $p = .193$.

These results show a practice effect, in that participants were significantly faster in each subsequent block, across the three conditions.

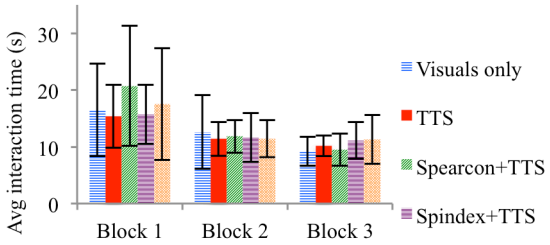


Figure 1: Average interaction time across the three blocks for the 5 conditions. Note that there was a significant difference between the three blocks across the conditions.

5.7.2. Accuracy

A two way repeated measures ANOVA (Huynh-Feldt corrections) was performed on the accuracy results and found no significant main effect of condition, $F(2.95, 61.85) = 0.67$, $p = .570$; no significant main effect of block, $F(1.40, 29.29) = 1.04$, $p = .342$; and no significant interaction, $F(5.51, 115.77) = 0.26$, $p = .944$. This means that accuracy was consistent in the song list experiment across conditions and blocks.

5.7.3. NASA-TLX

Perceived workload (NASA TLX) data are shown in Figure 2. A one way repeated measures ANOVA (Huynh-Feldt corrections) on the NASA-TLX results showed a significant

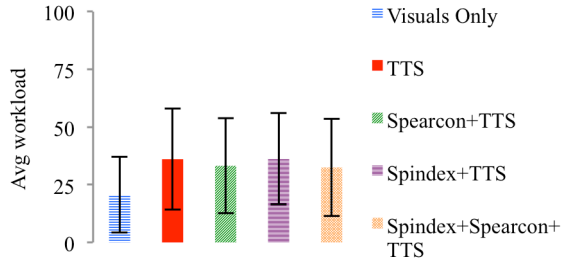


Figure 2: The average subjective workload as rated via NASA-TLX for each condition, averaged across blocks.

main effect of condition $F(2.87, 48.90) = 4.55$, $p = .008$. To determine where the differences between the conditions were taking place, a set of post-hoc analyses were performed via 10 paired t-tests with Bonferroni corrections (correcting alpha to .005). These analyses can be seen in Table 1. The tests revealed significantly less total subjective workload in the Visuals condition than the Spearcon+TTS condition, the Spindex+TTS condition, and the Spindex+Spearcon+TTS condition. No other differences were found to be significant. These results mean that participants reported higher total subjective workload in the Spearcon+TTS, Spindex+TTS, and Spindex+Spearcon+TTS conditions than they did in the Visuals condition.

5.7.4. Survey

The paired t-tests for the survey questions revealed no significant differences between the conditions in regards to preferences.

5.8. Discussion

The results of Study 1 showed participants were no faster or slower at finding the songs for any one condition, but were faster for each block on average across all conditions. These results suggest that the spindex and spearcon cues were able to convey information to the participants effectively with no training and that the extended number of cues was not a problem for participants. These results support the hypothesis that the spindex and spearcon cues could be used to navigate a long list in Mandarin, with no extended practice.

Block x Condition pairs	*t*	*df*	*p*
Visuals – TTS	2.97	17	.009
Visuals – Spearcon+TTS	3.51	17	.003*
Visuals – Spindex+TTS	4.14	17	.001*
Visuals – Spindex+Spearcon+TTS	3.77	17	.002*
TTS – Spearcon+TTS	0.24	17	.816
TTS – Spindex+TTS	0.53	17	.600
TTS – Spindex+Spearcon+TTS	0.42	17	.680
Spearcon+TTS – Spindex+TTS	0.64	17	.530
Spearcon+TTS – Spindex+Spearcon+TTS	0.55	17	.591
Spindex+TTS – Spindex+Spearcon+TTS	0.95	17	.354

Table 1: The paired t-test post hoc results done for TLX data in study 1. Note that * marks a significant difference.

The results also showed that participants had no significant difference in accuracy in finding the song across either condition or blocks. Again, this supports the hypothesis of the cues being able to be used effectively in Mandarin.

Although no time or accuracy differences were found, participants did report higher total workload in the Spearcon+TTS, Spindex+TTS, and Spindex+Spearcon+TTS conditions than they did in the Visuals condition. This hints that the subjective mental demand it took to complete the task was higher for the non-visual conditions, but this effect may decrease with practice.

文件	编辑	查看	收藏夹	工具	帮助
新建选项卡	剪切	工具栏	添加到收藏夹	删除浏览历史记录	Internet Explorer 帮助
重复打开选项卡	复制	浏览器栏	添加到收藏夹栏	InPrivate浏览	Internet Explorer 11中的新功能
新建窗口	黏贴	转到	将当前所有的网页添加到收藏夹	启用跟踪保护	联机支持
新建会话	全选	停止	整理收藏夹	ActiveX筛选	关于Internet Explorer
在沉浸式浏览器中打开	在此页上查找	刷新	Links	修复连接问题	
打开		缩放	Bing	重新打开上次浏览会话	
使用Notepad编辑		文字大小		将站点添加到“应用”视图	
保存		编码		查看下载	
另存为		样式		弹出窗口阻止程序	
关闭选项卡		插入光标预览		SmartScreen筛选器	
关闭所有标签		源		管理媒体许可证	
页面设置		安全报告		管理加载项	
打印		国际网站地址		兼容性视图设置	
打印预览		网页隐私策略		订阅此源	
发送		全屏		源发现	
导入和导出				Windows更新	
属性				性能仪表板	
退出				F12开发人员工具	
				OneNote Linked Notes	
				Send to OneNote	
				Internet选项	

Table 2: The Internet Explorer menu layout used in Study 2.

6. STUDY 2

6.1. Participants

A total of 23 participants (12 males and 11 females) with an average age of 20.6 (SD = 2.3) from the same university in the United States took part in the study. Again, only native speakers of Mandarin were recruited for the study and all participants were required to have normal or corrected to normal vision and hearing. Participants reported speaking Mandarin for an average of 20.3 years (SD=2.4) and writing the language for an average of 16.2 years (SD=4.9). The mean self-reported fluency ratings for Mandarin on a scale from 1-6 (6 being the highest fluency) was 5.9 (*SD*=0.3), and for writing it was 5.7 (*SD*=1.1). All participants signed informed consent and provided demographic information, and completed a few questions regarding their knowledge of Mandarin to ensure an equivalent minimum level of Mandarin expertise.

6.2. Apparatus

The apparatus for Study 2 was the same used in Study 1.

6.3. Menu structures

The menu system for Study 2 was based on the set of menu options available on Internet Explorer version 11, with one additional option ("Close all tabs") to create an even set of blocks to randomize within. This caused the structure of the menu out to have 69 menu items, under 6 tabs. The structure and items for the menu can be seen in Table 2.

6.4. Procedure

As in Study 1, participants were asked to confirm they met the study criteria and then read through and signed an informed consent form upon arrival. Participants were first introduced to the structure of the menu by looking and interacting with it visually. No audio was provided as they did this. During this time participants used the arrow keys to move left, right, up, and down, as they would during the study. They were also shown where the target item would be displayed on the screen once the study began. Following this orientation, participants started the condition blocks; these consisted of a training phase of 5 random item selections followed by 23 item selections (one for each available menu item) for each condition. The three conditions in the main menu blocks were Visual-only, TTS, and Spearcon+TTS; the order of conditions was randomized via a Latin square. Each condition was completed a total of 3 times, resulting in 9 total blocks. After each condition, participants completed the NASA-TLX to measure subjective workload.

Following the completion of all blocks of menu search tasks, participants completed a demographics survey and the same preferences questions as were given in Study 1.

6.5. Analysis

As in Study 1, data for interaction (search) time, accuracy, and TLX scores were all collected for each block. All of these data were analyzed via within subject repeated measures ANOVA with Huynh-Feldt corrections for sphericity. The survey data were analyzed via paired t-tests.

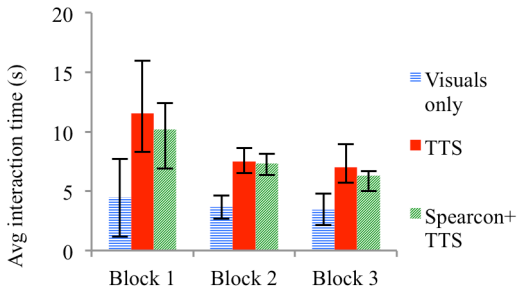


Figure 3: Average interaction time (in seconds) across the 3 blocks and 3 conditions for the IE menu task.

6.6. Results

6.6.1. Interaction time

The data for interaction (search) time across blocks and conditions are shown in Figure 3. A two way repeated measures ANOVA (Huynh-Feldt corrections) was done for interaction time, and revealed a significant main effect of condition $F(1.97, 37.47) = 90.80$, $p < .001$; a significant main effect of block, $F(1.30, 24.75) = 61.26$, $p < .001$; and a significant interaction, $F(1.93, 36.70) = 4.84$, $p = .014$.

To determine the differences in interaction time between conditions, post-hoc comparisons of 3 t-tests with Bonferroni corrections (alpha = .0166) were performed. The post-hocs revealed no significant difference between Spearcon+TTS and TTS, $t(21) = 1.21$, $p = .240$; but did show a significant difference between Spearcon+TTS and Visuals, $t(21) = 11.06$, $p < .001$; and for TTS and Visuals, $t(21) = 12.56$, $p < .001$. These results show that participants were significantly faster during the Visuals condition than both the Spearcon+TTS and the TTS conditions.

To determine the differences in interaction time between blocks, post-hoc comparisons of 3 t-tests with Bonferroni corrections (alpha = .0166) were performed. The post-hocs revealed significant differences between Blocks 1 and 2, $t(21) = 7.69$, $p < .001$; Blocks 1 and 3, $t(20) = 9.25$, $p < .001$; and Blocks 2 and 3, $t(20) = 4.20$, $p < .001$. These results reflect a practice effect, in that participants were significantly faster in each subsequent block.

To determine what interactions between condition and block were happening in the data, two sets of post-hoc comparisons of 9 t-tests (Bonferroni corrections; alpha = .0056) were performed. The first set of post-hocs was done to look at the interaction time differences between conditions within each block. These analyses can be seen in Table 3. The analyses revealed that the Visuals condition was significantly faster than either the Spearcon+TTS condition or the TTS condition within each block.

The second set of post-hocs was done to investigate the interaction time differences between blocks in each condition. The analyses showed that there were significant differences within the Spearcon-TTS and TTS conditions for each block. This means that participants got faster at using the Spearcon-TTS and TTS auditory cues as they progressed through the three blocks, which could be an argument for them learning how to use the cues more efficiently. No such learning effect was seen for the Visual-only condition, suggesting that it was not the learning of the menu that sped up the participant's interactions. The analysis data can be seen in Table 4.

Condition x Block pairs	t	df	p
Block 1 Spearcon+TTS – Block 1 TTS	1.05	21	.307
Block 1 Spearcon+TTS – Block 1 Visuals	6.61	21	< .001*
Block 1 TTS – Block 1 Visuals	6.36	21	< .001*
Block 2 Spearcon+TTS – Block 2 TTS	0.69	21	.495
Block 2 Spearcon+TTS – Block 2 Visuals	14.59	19	< .001*
Block 2 TTS – Block 2 Visuals	13.48	19	< .001*
Block 3 Spearcon+TTS – Block 3 TTS	1.74	20	.098
Block 3 Spearcon+TTS – Block 3 Visuals	9.61	20	< .001*
Block 3 TTS – Block 3 Visuals	8.34	20	< .001*

Table 3: Table of the post hoc analyses performed to investigate interaction time differences between conditions within each block for the IE menu task.

Block x Condition pairs	t	df	p
Block 1 Spearcon+TTS – Block 2 Spearcon+TTS	4.02	21	.001*
Block 1 Spearcon+TTS – Block 3 Spearcon+TTS	6.37	20	< .001*
Block 2 Spearcon+TTS – Block 3 Spearcon+TTS	3.89	20	.001*
Block 1 TTS – Block 2 TTS	4.15	21	< .001*
Block 1 TTS – Block 3 TTS	9.25	20	< .001*
Block 2 TTS – Block 3 TTS	15.32	20	< .001*
Block 1 Visuals – Block 2 Visuals	1.59	19	.128
Block 1 Visuals – Block 3 Visuals	1.96	20	.064
Block 2 Visuals – Block 3 Visuals	1.34	19	.197

Table 4: Table of the post hoc analyses performed to investigate interaction time differences between blocks in each condition

6.6.1. Accuracy

The data are shown in Figure 4. A two-way repeated measures ANOVA (Huynh-Feldt corrections) was done on the accuracy data, and found a main effect of block $F(1.18, 22.38) = 5.00$, $p = .031$. To determine the differences in accuracy between blocks post-hoc comparisons of 3 t-tests with Bonferroni corrections (alpha = .0166) were performed. The post-hocs revealed no significant difference between Blocks 1 and 2, $t(21) = 1.77$, $p = .091$; nor Blocks 1 and 3, $t(20) = 2.42$, $p = .025$; but did show a significant difference between Blocks 2 and 3 for the three conditions, $t(20) = 3.31$, $p = .003$.

The main ANOVA found no significant main effects of condition $F(1.36, 25.89) = .542$, $p = .522$, nor interactions $F(2.17, 41.18) = 1.52$, $p = .230$.

These results show a practice effect nearer the end of the study, in that participants were significantly more accurate across all three conditions in Block 3 than in Block 2.

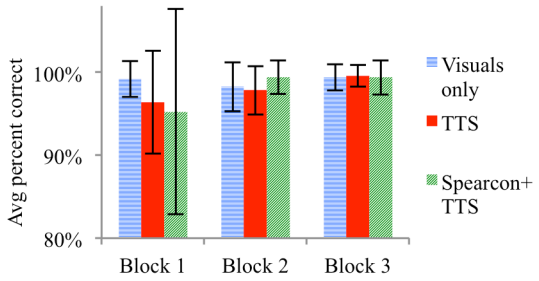


Figure 4: Average accuracy across conditions and blocks in the IE menu task.

6.6.2. NASA-TLX

A one way repeated measures ANOVA (Huynh-Feldt corrections) was done on the NASA-TLX results, and showed no significant main effects for condition $F(1.78, 35.66) = 1.33$, $p = .276$. This means that between the three conditions in the Internet Explorer study participants did not rate the conditions as being any different from each other in perceived workload.

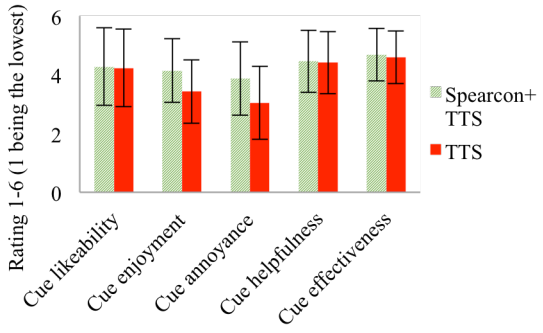


Figure 5: Average preference ratings across the spearcon+TTS and TTS cues in the IE menu study.

6.6.3. Survey

The survey data are shown in Figure 5. Paired samples t-tests on the survey data revealed that participants rated the Spearcon+TTS as more fun than TTS, $t(20) = 2.20$, $p = .040$. The analyses also revealed that participants considered Spearcon+TTS as more annoying than TTS, $t(21) = 2.41$, $p = .025$.

6.7. Discussion

In Study 2 it was found that participants were faster when using the visuals condition than either TTS or spearcon condition. This is expected, and suggests that the auditory-only cues, on the whole, slowed participants down; but the lack of any difference between spearcons and TTS for both speed and accuracy mean that the spearcons were effective at relaying information to the users. When investigating the learning effect on interaction (search) time, the results

showed that participants were getting faster in the two auditory conditions, but not the visuals condition. This suggests that the auditory conditions may have eventually been performed at the same speed as the visual conditions had participants had enough practice.

The results also showed, similarly to Study 1, that there was no significant difference in accuracy across the conditions, suggesting participants had similar ability in finding the correct menu item whether using spearcons, TTS, or visuals. The result of their learning was shown across all three conditions (Block 3 being more accurate than Block 2), but no other differences between conditions suggested similar performance across each condition.

Finally, NASA TLX results showed that participants had no perceived workload difference between the three conditions, meaning that the auditory cues were not subjectively harder to use than the visuals.

7. Overall Discussion

In sum, the results from the two studies suggest that spearcons and spindex cues are effective for use in Mandarin. The lack of any significant differences in accuracy across the Spindex or Spearcon cues and the TTS or visuals conditions suggests that participants were able to use the cues effectively to choose the required items. While time differences were seen between the advanced auditory cues and other cues in the IE task, participants would be expected to get even faster with these cues as they have more practice.

These results are similar to those seen in studies using spearcons in English [10] and Korean [12]. In these studies participants were also able to quickly learn to use the cues to effectively complete the tasks. It was seen in both of these previous studies that after a time of practice participants were actually faster with the spearcon cues. While this was not found in the current study, participants had less trials with the cues in this study so people may be found to be faster with more practice with the Mandarin version of the cues. Another similarity with the current study to previous work was that as with the study don in Korean [12], participants found spearcons to be more fun to use than TTS alone, which suggests people may be willing to use them in the real world.

7.1. Application of Results

These results suggest that the use of spindex and spearcon auditory cues in Mandarin could be effective in visually demanding multi-tasking situations or times when visual displays are not available. Screen readers for blind individuals could use these types of cues in Mandarin and most likely other tonal languages. As has been shown to work in English [2][19] these cues may help drivers to more safely perform secondary tasks in the car such as finding a radio station, or completing other tasks while keeping their eyes on the road.

7.2. Limitations

There were a few factors in the research that some might consider limitations including the use of college students and those who spoke English in addition to Mandarin. However, the use of students or their knowledge of English should not change the participants' abilities to perform the task as a Mandarin speaker. What should be considered is the performance with the auditory cues and the workload associated with them, as compared to the Visuals-only condition. The higher workload reported by the participants could be considered in applications of the work; however, having more practice with the cues would be expected to decrease this workload difference.

8. Conclusions

The results of this research suggest that Mandarin spearcons can work across an extended vocabulary and in multiple settings with no extensive training needed. In addition, the research suggests that spindex in Mandarin can help users move through a list effectively. This implies that these types of auditory cues can be used extensively in even more languages than previously known, and provide a suggestion that the cues can work in other tonal languages as well.

9. ACKNOWLEDGMENT

Portions of the work were supported by National Science Foundation Graduate Research Fellowships (DGE-1148903, DGE-1650044) as well as additional grant funding from the NSF and from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR). We would like to thank Grace Li and Alexis Wilkinson for collecting data for these studies.

10. REFERENCES

[1] Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., & Baudisch, P. (2007, April). Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1395-1404). ACM.

[2] Gable, T. M., Walker, B. N., Moses, H. R., & Chitloor, R. D. (2013, October). Advanced auditory cues on mobile phones help keep drivers' eyes on the road. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 66-73). ACM.

[3] Yin, M., & Zhai, S. (2006, April). The benefits of augmenting telephone voice menu navigation with visual browsing and search. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 319-328). ACM.

[4] Brumby, D. P., Davies, S. C., Janssen, C. P., & Grace, J. J. (2011, May). Fast or safe?: how performance objectives determine modality output choices while interacting on the move. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 473-482). ACM.

[5] Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. Human-computer interaction, 2(2), 167-177.

[6] Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. Human–Computer Interaction, 4(1), 11-44.

[7] Ferati, M., Pfaff, M. S., Mannheimer, S., & Bolchini, D. (2012). Audemes at work: Investigating features of non-speech sounds to maximize content recognition.

International Journal of Human-Computer Studies, 70(12), 936-966.

[8] Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D. K., Dingler, T., & Jeon, M. (2013). Spearcons (speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(1), 157-182.

[9] Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus.

[10] Palladino, D. K., & Walker, B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons.

[11] Jeon, M., & Walker, B. N. (2009, October). "Spindex": Accelerated Initial Speech Sounds Improve Navigation Performance in Auditory Menus. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 53, No. 17, pp. 1081-1085). SAGE Publications.

[12] Suh, H., Jeon, M., & Walker, B. N. (2012, September). Spearcons improve navigation performance and perceived speediness in Korean auditory menus. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 1361-1365). SAGE Publications.

[13] Larsson, P., & Niemand, M. (2015). Using sound to reduce visual distraction from in-vehicle human–machine interfaces. Traffic injury prevention, 16(sup1), S25-S30.

[14] Wersényi, G. (2010). Auditory representations of a graphical user interface for a better human-computer interaction. In Auditory Display (pp. 80-102). Springer Berlin Heidelberg.

[15] Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu navigation with in-vehicle technologies: auditory menu cues improve dual task performance, preference, and workload. *International Journal of Human-Computer Interaction*, *31*(1), 1-16.

[16] Davison, B. K., & Walker, B. N. (2008). AudioPlusWidgets: Bringing sound to software widgets and interface components. Proceedings of ICAD2008, Paris, France.

[17] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload, 1*(3), 139-183.

[18] SoX - Sound eXchange [Computer Software]. (2015). Retrieved from sox.sourceforge.net.

[19] Hussain, I., Chen, L., Mirza, H. T., Wang, L., Chen, G., & Memon, I. (2016). Chinese-Based Spearcons: Improving Pedestrian Navigation Performance in Eyes-Free Environment. International Journal of Human-Computer Interaction, 32(6), 460-469.

HOW DO PEOPLE *THINK* THEY REMEMBER MELODIES AND TIMBRES? PHENOMENOLOGICAL REPORTS OF MEMORY FOR NONVERBAL SOUNDS

Michael A. Nees, Joanna Harris, & Peri Leong

Lafayette College,
Department of Psychology,
Oechsle Hall, Easton, PA 18042, USA
[neesm][harrisje][leongp]@lafayette.edu

ABSTRACT

Memory for nonverbal sounds such as those used in sonifications has been recognized as a priority for cognitive-perceptual research in the field of auditory display. Yet memory processes for nonverbal sounds are not well understood, and existing theory and research have not provided a consensus on a mechanism of memory for nonverbal sounds. We report a new analysis of a qualitative question that asked participants to report the strategy they used to retain nonverbal sounds—both melodies and sounds discriminable primarily by timbre. The question was originally posed as part of the debriefing procedure for three separate memory experiments whose primary findings are reported elsewhere. Results of this new analysis suggested that auditory memory strategies—remembering acoustic properties of sounds—were common across both types of sounds but were more commonly reported for remembering melodies. Motor strategies were also more frequently reported for remembering melodies. Both verbal labeling of sounds and associative strategies—linking the sounds to existing information in memory—were more commonly reported as strategies for remembering sounds discriminable primarily by timbre. Implications for theory and future research are discussed.

1. INTRODUCTION

Kramer [1] identified the "absence of persistence" as a potential weakness of auditory displays. Unlike their visual counterparts, auditory displays are transient—a quality that may increase demands on memory, particularly if comparisons of sounds over time are required to accomplish tasks with auditory displays. Accordingly, the seminal Sonification Report [2] identified studies of memory for sounds as a priority in a research agenda for perceptual and cognitive scientists working in the field of auditory display. More specifically, Flowers [3] noted that sonifications should be designed to accommodate the limitations of auditory memory processes, including sensory memory and working memory.

Nearly two decades later, important research questions remain unresolved regarding memory for sounds—especially nonverbal sounds such as those used in sonifications. Influential theories of attentional and memory processes (e.g., [4], [5]) have little to say about cognitive processes for nonverbal sounds. This omission may result, at least in part, from the lack of certainty about basic representations of nonspeech sounds in memory.

https://doi.org/10.21785/icad2017.068

1.1. Auditory Sensory Memory and Working Memory

There is relatively wide agreement that sounds persist in memory for a brief period following stimulation. This phenomenon has been called *auditory sensory memory* or *echoic memory* in the literature. Auditory sensory memory preserves a high-fidelity mental representation of sounds in an auditory format—that is, for the duration of auditory sensory memory, sounds are accessible in memory as sounds per se. Further, auditory sensory memory processes do not seem to require attentional resources. The precise duration of auditory sensory memory is not currently known. Estimates have ranged widely—from a second or two in some studies up to 30 seconds or longer in others (for a review, see [6]).

Unlike the brief, passive auditory sensory store, *working memory* entails the active maintenance, rehearsal, and processing of information, and this active processing can sustain information in working memory indefinitely (see [6]–[8]). Verbal working memory, including auditory language, is widely believed to involve a subvocal (i.e., silent or covert) articulatory mechanism [4]. A miscellany of theoretical perspectives has emerged on the working memory rehearsal mechanism for nonverbal sounds. Baddeley and Logie suggested that [9] the working memory mechanism for verbal and nonverbal sounds might be one and the same. Berz [10] proposed a musical working memory mechanism that functioned with independence from verbal working memory processes. Attention-based rehearsal has been speculated to play a role in memory for timbre [11]—a property of sounds that some have purported is impossible to rehearse via subvocal articulation [12]. Yet other researchers have endorsed the perspective that nonverbal sounds cannot be rehearsed and are instead remembered "automatically" [13], although evidence to the contrary has been presented (e.g., [14]).

1.2. Encoding Strategies

Divergent hypotheses regarding the mechanism of working memory for nonverbal sounds may persist in part due to variability in the encoding strategies people use to remember sounds. People exhibit individual differences in the encoding strategies they use to remember perceptual stimuli, and these differences can be observed in behavioral studies of task performance. Macleod, Hunt, and Mathews [15], for example, exposed two distinct strategies used by participants in sentence-picture verification tasks. These tasks require participants to encode a sentence or phrase in memory (e.g., "plus is above star."). The time required to encode the stimulus is recorded as a dependent variable—comprehension

time. Following encoding, participants view a picture that either matches (e.g., a plus above a star) or does not match (e.g., a star above a plus) the state described in the sentence. Participants make a speeded response to indicate whether the sentence and picture matched, and the time required to respond is recorded as another dependent variable—verification time. Macleod et al. examined individual patterns of comprehension times and verification times and identified two different strategies for encoding the sentence. Participants who used a verbal encoding strategy remembered the sentence as words. Their characteristic pattern of task performance was short comprehension times and relatively longer verification times—presumably because they had to transform the sentence from its verbal memory code into a pictorial code for comparison during verification. Participants who used a visual imagery strategy exhibited long comprehension times—presumably because they were transforming the sentence to a picture during the comprehension phase of trials. Their verifications time were faster, however, as they could quickly compare their pictorial mental image to the picture stimulus. The individual difference that determined the strategy adopted seemed to be spatial ability; participants with higher scores on a psychometric test of spatial ability tended to adopt a visual imagery strategy. A later study [16] showed that participants could be trained to successfully adopt either strategy, regardless of their preferred default strategy.

1.3. Multiple Encoding Strategies for Melodies

Similarly, research on nonverbal sounds—melodies in particular—has suggested that people use multiple encoding strategies to remember tones. Zatorre and Beckett [17] reported that people with absolute pitch encode notes using note names (a verbal encoding strategy), motor codes (e.g., how to produce the note on a musical instrument), auditory codes (i.e., remembering sounds as sounds per se), and visual imagery (e.g., remembering the note by creating a mental image of its representation on a musical staff). Mikumo's (e.g., [18]) research suggested that these four codes may be widely available as encoding strategies, even for listeners who do not possess absolute pitch.

Nees and colleagues [19], [20] extended the sentence-picture verification task to include sound stimuli—brief, two-note nonspeech sounds that were like sonifications. Nees and Walker [19] demonstrated that tonal stimuli could be encoded as sounds (i.e., an auditory imagery strategy), as visuospatial images (i.e., a visual imagery strategy), or as words (a verbal strategy) depending upon the study instructions (also see [21]). The same study suggested that—for sounds only—the auditory sensory memory trace lingered for at least 3 seconds even when the stimulus had been recoded (e.g., to a verbal or visuospatial representation) in working memory.

1.4. Memory for Timbre

Although a good deal of research has suggested that a variety of encoding strategies can be used to remember melodies and pitched stimuli, the evidence regarding encoding strategies for sounds that are discriminable primarily by timbre is relatively scant. Crowder [22] provided evidence that timbre is stored in memory and can be imagined in the absence of a stimulus, presumably by reinstating representations of timbre held in long term memory. Crowder suggested that, to explain encoding of timbre, "an appeal to sensory rather than motor imagery is justified" (pp. 478) on the grounds that people cannot physically reproduce timbral differences in sounds. Our inability to physically reproduce sounds was assumed but not tested by Crowder. If correct, this would represent a difference between pitch and timbre memory, because pitch information can be rehearsed using a humming or singing encoding strategy.

Golubock and Janata [23] concluded that memory capacity for timbre was lower than verbal working memory capacity, as verbal items allowed access to both articulatory rehearsal and long-term associations. The specific auditory working memory capacity of sounds discriminable primarily by timbre was found to be just one to two items, in comparison to the typical verbal working memory limit of three to five items. Studies also have shown that the working memory capacity of sounds varying on timbre increases with more diverse stimuli [11], [23]—a finding which suggested that confusability among items with similar timbres could be a source of difficulty in memory for timbre.

1.5. The Current Study

Despite awareness in the auditory display community of the importance of memory for sounds, many questions remain regarding how people remember sounds—especially nonverbal sounds like those used in sonifications. The current study examined subjective reports of strategies used to encode melodies and timbre. Nees and Walker [24] previously reported a study on subjective reports of strategy use during data analysis tasks (such as point estimation) with auditory graphs—a type of sonification that uses pitch changes in time to convey information (for a review, see [25]). They found that, although verbal, visuospatial, auditory, and motor encoding strategies were reported, of the four, verbal strategies were reported most frequently. Further, each of these strategies was reported less frequently than other strategies (e.g., counting in time to the auditory graph and using the auditory graph's contextual auditory cues) that were specific to the data analysis task (rather than representative of a general encoding strategy for sounds). The general approach used by Nees and Walker was adopted here, although the tasks of the current analyses reflect demands that were more purely based in memory than the tasks used by Nees and Walker.

The data analyzed here were collected as part of a debriefing procedure across three previous experiments whose primary results are reported elsewhere (see below). Participants gave an open-ended response indicating the strategy they had used to retain either melodies or the timbre of sounds during an 8 second retention interval. This exploratory investigation sought to examine participants' subjective impressions of how they rehearsed and retained nonverbal sounds in memory during these experiments, and also to examine potential differences in these impressions when the target stimuli were pitched melodies as compared to sounds discriminable by timbre only.

2. METHOD

The data reported here were collected as part of a qualitative post-experimental follow-up question in three separate experiments. The primary results of these experiments are reported elsewhere. The first two experiments (reported in [26]) examined memory for melodies, and the third experiment [27] examined memory for timbre. Although the focus of the current report is the analysis of a single qualitative follow-up question (see section 2.3), for context a brief description of the methods of each experiment follows.

2.1. Memory for Melodies Experiments

Nees, Corrini, Leong, and Harris [26] conducted a pair of studies that were designed to examine both articulatory processes and attentional processes as potential mechanisms of rehearsal of melodies in working memory. The stimuli and task were modeled after the task used by Schendel and Palmer [28]. Participants heard brief four-note melodies, with each note randomly selected from the octave ranging from D4 to E5. Notes were created in the MIDI piano timbre and were 500 ms in duration, including 10 ms onset and offset ramps; each melody was 2000 ms in total duration. On a trial, participants heard a standard melody. Following an 8 second delay, participants heard a second probe melody and indicated whether the probe melody was the same as the initial standard melody. Half of the trials in the study featured the same melody, and half featured different melodies. Different melodies were altered such that either the second or third note in the melody was either raised or lowered by one note. Across four blocks, participants experienced a set of distractor conditions intended to suppress articulatory rehearsal (reading solved math problems aloud), attentional rehearsal (solving math problems and responding manually instead of vocally), or both (reading and solving math problems aloud). In the first experiment, articulatory rehearsal suppression was spoken aloud, whereas in the second experiment, it was silently voiced. Results of both experiments showed that articulatory rehearsal suppression interfered with memory for melodies, whereas interference with attentional rehearsal did not. The studies were interpreted to indicate that articulatory rehearsal plays a role in working memory for melodies.

2.2. Memory for Timbre Experiment

Nees, Leong, and Harris [27] conducted an experiment to follow-up the melodies experiment described above. The stimuli for this experiment were 8 sounds culled from the stimulus set reported in Golubock and Janata ([23], Experiment 1, pp. 402). These synthesized stimuli were designed to be abstract and discriminable by timbre. Golubock and Janata began with a sound that had a fundamental frequency of 341 Hz and 19 harmonics. Next, they systematically altered the attack, spectral centroid, and spectral flux of the sound to create a set of stimuli that were placed in a three-dimensional (attack, centroid, flux) stimulus space. Finally, they conducted listening tests and confirmed that the sounds at the corners of their design space were discriminable. We used the 8 corner stimuli from their stimulus space in our study. Although the sounds were discriminable, they were designed such that discrimination would not be possible using cues related to pitch or existing associations with instruments or real-world sound sources.

The Nees et al. timbre experiment [27] followed the same basic procedure as the melodies experiments described above, except that participants heard only one abstract sound and one probe. The primary timbre experiment found no difference across the interference conditions. Participants exhibited much higher d' (sensitivity) scores in this experiment as compared to the melodies experiments above. We concluded that the task exhibited a form of ceiling effect, and follow-up studies are underway with longer sequences of abstract sounds.

2.3. The Current Study: Reports of Strategy Use

2.3.1. Participants

There were 40 participants in the first melody study (30 females, *M* age = 19.58 years, *SD* = 1.22), 36 participants in the second melody study (28 females, *M* age = 19.53 years, *SD* = 0.91), and 44 participants (34 females, *M* age = 19.41 years, *SD* = 0.87) in the timbre study. Data from one participant in the first melody study were not usable, so the final analysis reported here involved a total sample of $N = 119$ cases. In all 3 studies, participants rated their musical ability on a scale from 1 ("I have no musical ability at all") to 7 ("I am a professional musician"). The ratings across the two melodies studies and the timbre study were $M = 2.85$ ($SD = 1.44$, $mdn = 3.00$, mode = 2), $M = 2.78$ ($SD = 1.44$, $mdn = 3.00$, mode = 1), and $M = 2.68$ ($SD = 1.54$, $mdn = 2.00$, mode = 1), respectively, which suggested that participants were overwhelmingly non-musicians.

2.3.2. Strategy Use Query

At the end of each of the studies described in 2.1 and 2.2, participants responded to the following question:

"Please briefly describe the strategy you used to remember the [sounds]. If you used different strategies for different parts of the study, describe the different strategies."

The question was open-ended. Participants typed their answers into a text box on a Qualtrics questionnaire. They could type as much or as little as they wished.

2.4. Coding Rubric

The coding rubric for this analysis was adapted from the coding scheme used in Nees and Walker [24]. Nees and Walker's task and stimuli were different from the tasks and stimuli used here, so some modifications to their coding scheme were required. The current rubric adopted five of the strategies described in their rubric directly (verbal, visuospatial, motor, auditory-musical, and differential strategies). Three of the strategies they described were specific to their tasks and stimuli (context, counting, and arithmetic strategies), and we replaced those categories with strategies that were relevant to our tasks and stimuli (i.e., the associative and distraction avoidance strategy categories, and also the no strategy category). The rubric's operational definitions of statements that fell into each strategy category are described below.

2.4.1. Verbal Strategy

A response was coded as indicating a verbal encoding strategy if the participant reported naming sounds, labeling sounds (e.g., any mention of musical note names), comparing sounds to one another using assigned verbal labels (such as "high" and "low"), or mentioning specific sounds by an assigned name. This strategy included any indication that the subject labeled specific sounds with a verbal (i.e., linguistic) tag.

2.4.2. Visuospatial Strategy

A response was coded as indicating a visuospatial encoding strategy if the participant reported that she mentally drew a

picture, used visual imagery, or created a graph her head. Responses that suggested features characteristic of a visual image (slope, line, top, bottom etc.) were also included.

2.4.3. Motor Strategy

A response was coded as indicating a motor encoding strategy if the participant reported a non-articulatory movement strategy (meaning use of the hands or feet, etc.) to assist memory. This strategy included counting on fingers, tapping, moving physically with the mouse or finger or "drawing" on the desk with a finger to remember contour.

2.4.4. Auditory-musical Strategy

A response was coded as indicating an auditory-musical encoding strategy if the participant reported humming, whistling, vocalizing, or imitating, either overtly or covertly, the melody, pitch, or timbre of any part of the stimulus. This strategy included an indication that the participant was attempting to maintain some isomorphic representation of the sound or replay the sound stimulus. Any mention of the participant "hearing" the sound, recording the sound in her mind, or "focusing" on specific notes, etc., also was included in this category.

2.4.5. Associative Strategy

A response was coded as indicating an associative encoding strategy if the participant reported making connections with any information that was already learned. This strategy included associating sounds with existing sounds or other information in long-term memory and use of mnemonics.

2.4.6. Differential Strategy

A response was coded as indicating a differential encoding strategy if the participant reported differences in strategy use based on stimulus differences or experimental conditions. This strategy included responses that offered any indication that the strategy switched over the course of the experiment.

2.4.7. Distraction Avoidance Strategy

A response was coded as indicating a distraction avoidance encoding strategy if the participant reported trying to ignore the interference tasks or to control attention to minimize distraction and maximize focus on the sounds. This strategy included responses indicating that the participant closed her eyes to avoid visual distractions, etc.

2.4.8. No Strategy or No Knowledge of Strategy

A response was coded as indicating no encoding strategy or no knowledge of the strategy if the participant reported having no knowledge of her strategy or indicated that she used no strategy at all.

2.5. Coding Procedure

The coding procedure involved three steps. First, all authors examined all participant responses for fit with the rubric used by Nees and Walker [24] without performing any formal ratings. Second, in a meeting, all three authors discussed the Nees and Walker rubric and its rating criteria. During the meeting, we arrived at a consensus about the applicability of each strategy to our data and adapted some categories of the rubric as described above. Finally, two raters (the second and third authors) independently coded the presence or absence of each strategy in each response across all cases. Coding of strategies was not mutually exclusive—a given response could be categorized as meeting the criteria for more than one strategy.

3. RESULTS

3.1. Length of Responses

All responses were coded for length using the word count feature of Microsoft Word. The mean length of responses was $M = 25.68$ words ($SD = 25.37$).

3.2. Inter-rater Agreement

The coding scheme was checked for inter-rater agreement among the two raters using a percent agreement measure. Percent agreement was calculated to account for the number of discrepant ratings (i.e. that showed lack of agreement in coding) across all coded categories for the $N = 119$ cases as shown in Equation 1.

$$Percent\ agreement = \frac{119 - total\ discrepancies}{119} \quad (1)$$

Percent agreement was high across the raters (ranging from 92% to 98% agreement for the strategies); thus, no revision to the coding scheme or re-rating was deemed necessary. The final determination of how to code the small number of discrepant ratings was settled by discussion among the raters to arrive at consensus.

3.3. Overall Use of Strategies Across All Cases

Table 1 shows the overall reported use of each strategy across all cases. Notable results here included the high percent of respondents that used an auditory-musical strategy and the low percent of participants that reported having no strategy or no knowledge of their strategy. Most participants, then, apparently could access and report upon a specific approach they used to remember sounds. Perhaps not surprisingly, most reported using a strategy that involved trying to remember the sounds as sounds per se. Verbatim representative examples of the types of statements that fell into each strategy category are given below.

Strategy	Overall Percent Reported (N = 119)
Verbal	18
Visuospatial	12
Motor	16
Auditory-musical	61
Associative	13
Differential	8
Distraction Avoidance	12
No strategy/ knowledge	5

Table 1. Percent of participants reporting using each strategy.

3.3.1. Verbal Strategy

Across all cases, 18 percent of participants reported using a verbal strategy. Representative examples of statements indicating a verbal encoding strategy are shown below.

"*I tried to think of/remember whether the noise was high, low, or in the middle…"*

"I'd say quick words to the pitch and anticipate the second round with those words."

3.3.2. Visuospatial Strategy

Across all cases, 12 percent of participants reported using a visuospatial strategy. Representative examples of statements indicating a visuospatial encoding strategy are shown below.

"I visualized the notes on a scale as if I was reading music."

"I counted the beats and graphed them in my head in almost a bar graph."

3.3.3. Motor Strategy

Across all cases, 16 percent of participants reported using a motor strategy. Representative examples of statements indicating a motor encoding strategy are shown below.

"I used my fingers to tap it on different locations on the mousepad. The locations corresponded to different pitches."

"Tapping out the music on my fingers."

"I would look up if the note was high and down if the note was low and followed the beat with my eyes."

3.3.4. Auditory-musical Strategy

Across all cases, 61 percent of participants reported using an auditory-musical strategy. Representative examples of statements indicating an auditory-musical encoding strategy are shown below.

"I sung it in my head during the pause."

"I just repeated the melodies in my head as I answered the problems."

3.3.5. Associative Strategy

Across all cases, 13 percent of participants reported using an associative strategy. Representative examples of statements indicating an associative encoding strategy are shown below.

"I associated the noises with different sound effects."

"Sounds reminded me of certain things. One sounded like an instrument being aggressively plucked, another like an angry pop up, another sounded like it got cut off, etc."

3.3.6. Differential Strategy

Across all cases, 8 percent of participants reported using a differential strategy. Representative examples of statements indicating a differential strategy are shown below.

"For the sections with the math problem, I would try to play out the melody like I was playing it on the piano. For the section without the math problems, I tried to visualize the melody on the staff then repeat it 3 times before it played again."

"During the portions with math I tried to read the math as fast as possible so I could keep trying to repeat the melody in my head with little interruption. During the portion without math I tried to repeat the melody and the high low beats as much as I could before the second melody played."

3.3.7. Distraction Avoidance Strategy

Across all cases, 12 percent of participants reported using a distraction avoidance strategy. Representative examples of statements indicating a distraction avoidance strategy are shown below.

"For the last part of the study with the 8 second gap in between each sound, I closed my eyes when listening to the sounds so that I would have no visual distractions."

"I tried to pay as little attention as possible to the math problems."

3.3.8. No Strategy or No Knowledge of Strategy

Across all cases, 5 percent of participants reported having no strategy. Representative examples of statements indicating no knowledge of strategy are shown below.

"I did not really have a strategy."

"I tried to listen carefully but I did not have a strategy."

3.4. Comparison of Strategy Use Across Melody and Timbre Stimuli

Chi-square tests (see [29], [30]) compared the proportions of respondents that used each strategy to remember melodies and timbres (see Footnote 1). The results of these analyses are presented in Table 2.

Results showed that participants were significantly more likely to report having used motor and auditory-musical strategies to remember melodies as compared to timbre. For

timbre, participants were significantly more likely to report having used verbal or associative encoding strategies.

Strategy	Melody Percent (N = 75)	Timbre Percent (N = 44)	χ^2 (1 *df*)	*p*
Verbal	9	32	10.07	.002*
Visuospatial	16	5	3.17	.08
Motor	23	5	6.52	.01*
Auditory-musical	71	45	7.84	.005*
Associative	3	32	19.53	<.001*
Differential	9	7	0.15	.70
Distraction Avoidance	12	11	0.03	.87
No strategy/ knowledge	7	2	1.41	.24

Table 2. Percent of use of each strategy reported in the melody studies and timbre study and associated chi-square test results. Asterisks indicate $p < .05$.

4. DISCUSSION

Results showed that overall, most participants could report information about the strategy they believed they had used to remember sounds during a retention interval. Perhaps not surprisingly, an auditory-musical strategy was the most frequently reported strategy.

Some notable differences in strategy use were reported for remembering melodies versus timbre. Significantly more participants reported using a motor strategy to remember melodies as compared to timbre. This suggested that the pitch information in melodies may map more readily to motor responses (finger tapping, etc.). Research has indicated that higher pitch is associated with higher locations in space [31]. As such, people may be able to use motor encoding that represents pitch in physical space to supplement memory for melodies (see section 3.3.3). Previous research [17] has suggested that musicians may encode pitch using motor codes that involve production of the note on their instruments. The current study's results suggested that non-musicians might use more generic or rudimentary motor encoding strategies that link the pitch of tones to locations in space with motor programs.

Although participants reported high use of an auditory-musical encoding strategy for both melodies and timbre, this strategy was reported significantly more frequently in the melody studies. The sounds in the timbre study were abstract, and, although they were discriminable, some sounded perceptually similar. Perhaps these features discouraged or suppressed some participants' attempts to remember the timbral sounds as sounds per se and thus prompted them to attempt different encoding strategies.

Two strategies—verbal labeling and an associative strategy—were reported significantly more frequently among participants in the timbre study as compared to the melody study. Both strategies involve a version of elaboration—linking the abstract timbral stimuli to other information (by verbal labelling or associating the abstract sounds with existing knowledge) in an effort to aid memory. Elaborative strategies can aid in memory tasks (see [32]), especially when the bottom-up stimulus has no inherent meaning, as was the case here.

Taken together, the data presented here suggest that people use a robust variety of strategies to remember nonverbal sounds, and the most frequently used strategy was an auditory strategy that attempted to remember the sounds as sounds per se. Our findings suggested that abstract sounds may be more difficult than melodies to remember based on their acoustic properties, thus some people may try to encode them by linking them with information already present in their memory repertoire (e.g., verbal labels or names, other known sounds, etc.).

4.1. Limitations

Our approach in this study has limitations that should be considered when interpreting our findings. First, our data were limited to those strategies that participants chose to report. A participant's failure to mention using a particular strategy does not ensure that the participant did not use that strategy; instead, she may have used the strategy but chosen not to report it. The open-ended format of the question allowed participants to type as little as they wished. Some participants may have offered incomplete reports regarding their strategy use. Since the strategy question was asked at the end of the experiment, it also required participants to remember their strategies. It is possible that some participants had forgotten important aspects of their encoding strategies by the end of the study.

Further, participants can only report accurately on their strategy use to the extent that their memory processes were consciously accessible to them. Although the assumption that encoding strategies are accessible, reportable, and under voluntary command has been accepted in various research paradigms in cognitive psychology (e.g., [33], [34]), this assumption has also been challenged as an invalid appeal to introspection [35]. Our position on this matter lies somewhere in between these two perspectives. We acknowledge that the strategies reported by our participants cannot necessarily be assumed to accurately capture their actual mental processes. We also suspect, however, that in many cases the subjective reports do provide some information about those mental processes and point to interesting areas for future research with objective, quantitative approaches to strategy measurement. Further, we find the phenomenon of subjective impressions of strategy use to be one that is of interest in and of itself. That is, the phenomenological aspect of perception—"*what it is like to perceive"* a stimulus—has garnered serious attention as a topic of inquiry unto itself (see [36], pp. 2), yet few reports have examined this topic with respect to memory for melodies and timbre.

4.2. Directions for Future Research

Research consistently has suggested that multiple strategies can be used to encode pitch information [17]–[19], [21]. Less is known about encoding strategies for timbre. Researchers (e.g., [22]) have suggested that timbre memory entails auditory-sensory coding that maintains some isomorphic and phenomenologically auditory (e.g., rather than motor) representation of the timbral features of sounds.

Still, little is known about the functionality of cognitive mechanism(s) of auditory encoding strategies that preserve the acoustic characteristics of sounds. In particular, it is unclear how purely auditory (nonverbal, etc.) representations are rehearsed in working memory. Some researchers (e.g., [13],

[22], [23]) seem to have adopted the perspective that working memory for timbre entails a version of auditory sensory memory (see [6]) that is protracted in duration, though the mechanism that permits the representation to endure in working memory following perception has not been specified. More research is needed to establish how auditory representations are rehearsed in working memory.

5. CONCLUSIONS

Auditory displays and sonifications that use pitch to convey information may invoke different memory strategies than those that use abstract properties of sounds that must be discriminated based on timbre. Whether the use of different strategies ultimately will be an advantage or a hindrance to the end user will depend upon the particular use scenario (task constraints, etc.) and display design, but the deployment of auditory displays should proceed with awareness that different types of sounds may engage different cognitive processes in memory.

6. FOOTNOTES

1. The calculator used to compute the Chi-square tests is available online at:
https://www.medcalc.org/calc/comparison_of_proportions.php

7. REFERENCES

[1] G. Kramer, "An introduction to auditory display," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed. Reading, MA: Addison Wesley, 1994, pp. 1–78.

[2] G. Kramer *et al.*, "The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display," 1999.

[3] J. H. Flowers, L. E. Whitwer, D. C. Grafel, and C. A. Kotan, "Sonification of daily weather records: Issues of perception, attention, and memory in design choices," in *Proceedings of the International Conference on Auditory Display,* Espoo, Finland, 2001, pp. 222–226.

[4] A. D. Baddeley, "Working memory: Theories, models, and controversies," *Annual Review of Psychology*, vol. 63, pp. 1–29, 2012.

[5] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, pp. 159–177, Jan. 2002.

[6] M. A. Nees, "Have We Forgotten Auditory Sensory Memory? Retention Intervals in Studies of Nonverbal Auditory Working Memory," *Frontiers in Psychology*, vol. 7, 2016.

[7] R. H. Logie and N. Cowan, "Perspectives on working memory: introduction to the special issue," *Memory & Cognition*, vol. 43, no. 3, pp. 315–324, 2015.

[8] A. Miyake and P. Shah, *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press, 1999.

[9] A. D. Baddeley and R. H. Logie, "Auditory imagery and working memory," in *Auditory Imagery*, D. Reisberg, Ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992, pp. 179–197.

[10] W. L. Berz, "Working Memory in Music: A Theoretical Model," *Music Perception: An Interdisciplinary Journal*, vol. 12, no. 3, pp. 353–364, Apr. 1995.

[11] K. Siedenburg and S. McAdams, "The role of long-term familiarity and attentional maintenance in short-term memory for timbre," *Memory*, vol. 25, no. 4, pp. 550-564.

[12] R. G. Crowder, "Auditory memory," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, S. McAdams and E. Bigand, Eds. New York, NY, US: Clarendon Press/Oxford University Press, 1993, pp. 113–145.

[13] L. Demany and C. Semal, "The role of memory in auditory perception," in *Auditory Perception of Sound Sources*, Springer, 2008, pp. 77–113.

[14] T. A. Keller, N. Cowan, and J. S. Saults, "Can auditory memory for tone pitch be rehearsed?," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 21, no. 3, pp. 635–635, 1995.

[15] C. M. Macleod, E. B. Hunt, and N. N. Mathews, "Individual differences in the verification of sentence—picture relationships," *Journal of Verbal Learning and Verbal Behavior*, vol. 17, no. 5, pp. 493–507, Oct. 1978.

[16] N. N. Mathews, E. B. Hunt, and C. M. MacLeod, "Strategy choice and strategy training in sentence-picture verification," *Journal of Verbal Learning and Verbal Behavior*, vol. 19, no. 5, pp. 531–548, 1980.

[17] R. J. Zatorre and C. Beckett, "Multiple coding strategies in the retention of musical tones by possessors of absolute pitch," *Memory & Cognition*, vol. 17, no. 5, pp. 582–589, 1989.

[18] M. Mikumo, "Multi-encoding for pitch information of tone sequences," *Japanese Psychological Research*, vol. 39, no. 4, pp. 300–311, 1997.

[19] M. A. Nees and B. N. Walker, "Flexibility of Working Memory Encoding in a Sentence-Picture-Sound Verification Task," *Journal of Cognitive Psychology*, vol. 25, no. 7, pp. 800–807, 2013.

[20] M. A. Nees and K. Best, "Modality and encoding strategy effects on a verification task with accelerated speech, visual text, and tones," in *Proceedings of the International Conference on Auditory Display*, Lodz, Poland, 2013, pp. 267–274.

[21] M. A. Nees and B. N. Walker, "Mental scanning of sonifications reveals flexible encoding of nonspeech sounds and a universal per-item scanning cost," *Acta Psychologica*, vol. 137, no. 3, pp. 309–317, 2011.

[22] R. G. Crowder, "Imagery for musical timbre," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 15, no. 3, pp. 472–478, 1989.

[23] J. L. Golubock and P. Janata, "Keeping timbre in mind: Working memory for complex sounds that can't be verbalized," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 39, no. 2, pp. 399–412, 2013.

[24] M. A. Nees and B. N. Walker, "Encoding and representation of information in auditory graphs: Descriptive reports of listener strategies for understanding data," in *Proceedings of the International Conference on Auditory Display,* Paris, France, 2008.

[25] M. A. Nees and B. N. Walker, "Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension," in *Proceedings of the International Conference on Auditory Display,* Montreal, Canada, 2007, pp. 266–273.

[26] M. A. Nees, Corrini, Ellen, Leong, Peri, and Harris, Joanna, "Maintenance of Memory for Melodies:

Articulation or Attentional Refreshing?," *Psychonomic Bulletin & Review*, In Press.
[27] M. A. Nees, Leong, Peri, and Harris, Joanna, "Mechanisms of working memory for timbre," Unpublished data.
[28] Z. A. Schendel and C. Palmer, "Suppression effects on musical and verbal memory," *Memory & Cognition*, vol. 35, no. 4, pp. 640–650, Jun. 2007.
[29] J. T. E. Richardson, "The analysis of 2 × 2 contingency tables—Yet again," *Statistics in Medicine*, vol. 30, no. 8, pp. 890–890, Apr. 2011.
[30] I. Campbell, "Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations," *Statistics in Medicine*, vol. 26, no. 19, pp. 3661–3675, Aug. 2007.
[31] E. Ben-Artzi and L. E. Marks, "Visual-auditory interaction in speeded classification: Role of stimulus difference," *Perception & Psychophysics*, vol. 57, no. 8, pp. 1151–1162, 1995.
[32] F. I. M. Craik and R. S. Lockhart, "Levels of processing: A framework for memory research," *Journal of Verbal Learning and Verbal Behavior*, vol. 11, no. 6, pp. 671–684, Dec. 1972.
[33] J. Marquer and M. Pereira, "Reaction times in the study of strategies in sentence–picture verification: A reconsideration," *The Quarterly Journal of Experimental Psychology Section A*, vol. 42, no. 1, pp. 147–168, 1990.
[34] S. M. Kosslyn, *Image and Mind*. Cambridge, MA: Harvard University Press, 1980.
[35] J. S. B. Evans, "Verbal reports of cognitive strategies: A note on Marquer and Pereira," *The Quarterly Journal of Experimental Psychology*, vol. 42, no. 1, pp. 169–170, 1990.
[36] W. Fish and others, *Philosophy of Perception: A Contemporary Introduction*. Routledge, 2010.

8. ACKNOWLEDGMENTS

We thank Ellen Corrini for her contribution to data collection during the first memory for melodies experiment. Petr Janata graciously shared with us the stimuli from Golubock and Janata [23], and a subset of those sounds were the stimuli used in our memory for timbre experiment.

REFLECTIONS ON THE REPRESENTATION OF WOMEN IN THE INTERNATIONAL CONFERENCES ON AUDITORY DISPLAYS (ICAD)

Areti Andreopoulou

Laboratory of Music Acoustics and Technology (LabMAT)
University of Athens, Greece
areti.andreopoulou@gmail.com

Visda Goudarzi

Institute of Electronic Music and Acoustics (IEM)
Graz, Austria
goudarzi@iem.at

ABSTRACT

This paper investigates the representation of women researchers and artists in the conferences of the International Community for Auditory Display (ICAD). In the absence of an organized membership mechanism and / or publicly available records of conference attendees, this topic was approached through the study of publication and authorship patterns of female researchers in ICAD conferences. Temporal analysis showed that, even though there has been an increase in the number of publications co-authored by female researchers, the annual percentage of female authors remained in relatively unchanged levels (mean = 17.9%) throughout the history of ICAD conferences. This level, even though low, remains within the reported percentages of female representation in other communities with related disciplines, such as the International Computer Music Association (ICMA) and the Conferences of the International Society for Music Information Retrieval (ISMIR), and significantly higher than in more audio engineering-related communities, such as the Audio Engineering Society (AES).

1. INTRODUCTION

The International Community for Auditory Display (ICAD) is a highly interdisciplinary body focusing on the use of sound to perceptualize data. The foundations of this scientific discipline were laid during the first ICAD conference in 1992, organized by Gregory Krammer [1]. Since then, ICAD has been known for bringing together researchers from various research fields, including but not limited to music technology, acoustics, psycho-acoustics, music composition, computer science, and electrical engineering.

Despite the community's strong ties to areas related to the arts and science, ICAD-related research has been dominated by male researchers. This fact can be attributed to its equally strong ties to audio engineering and other technical fields. In response to this observation, over the past few years, ICAD conference organizers have started hosting women-related events, during which such issues are openly acknowledged and discussed. However, a more systematic approach, which would promote and encourage the involvement of female researchers in leadership roles, such as session, conference, and program chairs, reviewers and meta-reviewers, mentors, and Board of Directors members, is still missing. In an attempt to yield empirical data supporting the need for such initiatives, the authors have conducted analysis on the publication and authorship patterns of female researchers in ICAD conferences.

Over the past few years, several similar papers have been published about other related communities, such as the Audio Engineering Society (AES), the International Computer Music Association (ICMA) [2], and the Conferences of the International Society for Music Information Retrieval (ISMIR) [3]. Even though cross-study comparisons are not feasible due to differences in methodology and data analysis, each publication offers a valuable perspective on the topic. Hopefully, this increase in the number of studies revealing the under-representation of women scientists and artists in audio and music-related research-fields, will stimulate people's interest in this issue and lead to the necessary actions to resolve the situation.

2. RELATED WORK

In the majority of technical fields the number of female researchers and scholars is significantly smaller than that of their male colleagues. A report from the American Association of University Women states that women account for 17% of the high school students who take advanced placement exams in computer science, and for 28% of the undergraduate degrees awarded on related disciplines. Similar to computer science, in the fields electronic and computer music both the aesthetic path of music composition and the technical aspects of audio engineering and recording have been dominated by male researchers.

The early literature on 20^{th} century music undermined women's achievements in composition and performance [4]. Early music composition books referenced one to two female composers. It was not until the mid '90s that female electronic music composers such as Pauline Oliveros and Kaija Saariaho started to be included in textbooks, and gender-biased computer music advertising to be criticized [5]. An attempt to shed light on this situation was made by Mathew et al. [2], who attributed women's under-representation in fields related to computer music and audio to a) social and environmental factors [6], b) lack of role models [7], and c) issues related to work-life balance [8].

Women's under-representation extends beyond education, artistic creation, and engineering to scientific research. This was demonstrated in a study by Sugimoto et al.[9] who investigated the percentage of female authors in 5,483,841 articles published between 2008 and 2012 in the Web of Science, an online database

https://doi.org/10.21785/icad2017.031

of 27 million publications. Results indicated that only 30% of the authors were female. Hence, the gender gap in scientific research publications appears to be prominent in a much broader range of disciplines. In certain cases, this bias seems to affect how women's work is valued within a scientific community. A study of citation patterns based on gender in the field of International Relations [10] reported that papers authored by women were cited significantly less than those written by men. Nevertheless, it is worth mentioning here that a similar study on the scientific works of the audio-related ISMIR community revealed no such tendencies [11].

Previously published work on the representation of women scientists and artists in music and audio-related fields has found the rates of female members to be $< 20\%$ in the International Computer Music Association (ICMA) and $< 10\%$ in the Audio Engineering Society (AES) [2]. Similar analysis on the representation of female authors in the proceedings of the Conferences of the International Society for Music Information Retrieval (ISMIR) estimated the participation of female authors to vary between 10% and 20% [3, 11]. This paper will present an overview of the representation of female scientists and artists in the published proceedings of all ICAD conferences since 1994.

3. METHODOLOGY

3.1. Data collection and standardization

In the absence of either a membership mechanism or publicly available records of ICAD conference attendees, the investigation of the representation of women in the International Community of Auditory Displays was based exclusively on the infometrics of authors with published papers in ICAD conferences. ICAD proceedings, which are hosted by the SMARTech Repository service of the Georgia Tech Library, are publicly available under a Creative Commons Attribution 4.0 International License.

Records of papers and authors for all available conference-years between 1994 and 2016 were extracted from the SMARTech website and stored into spreadsheets. Data extraction was followed by a standardization procedure during which authors were cataloged in the following format: ["Last name", "First name" "Initials of middle names (if any)"]. For cases where only the initials of first names were provided, full names were manually retrieved through the search of public records on the Internet. Following that, the authors' gender was manually determined based on their first names. Additional information on the number of authors per publication and their affiliation was also logged.

3.2. Limitations

The authors acknowledge that the selected methodology for the study of women's representation in the International Community for Auditory Display suffers from certain limitations. First, a study of the representation of female authors in ICAD publications cannot fully reflect the involvement of women in the conferences, where their role could be that of an attendee, conference staff, or organizer, nor their involvement in the community in general. Nevertheless, in the absence of an official ICAD membership mechanism, published ICAD proceedings are the only reliable and publicly available source of information, to date.

In addition, the procedure followed to determine the authors' gender based on their first-names is error-prone for various reasons. First-names are not always gender specific across all languages. For cases where it was not possible to determine one's gender based on their name, a search on the Internet was performed using the author's full name and affiliation and a gender label was assigned based on photographs and / or online bios. Moreover, certain names can be exclusively attributed to a single gender in one language but be gender neutral or attributed to the opposite gender in other languages. Even though gender labels were assigned with caution and, to the extend possible, in accordance with the author's nationality, it is possible that some of the data have been mislabelled. Finally, this work uses a simplified binary gender distinction to refer to all ICAD authors. It is possible that some

[0]https://smartech.gatech.edu/

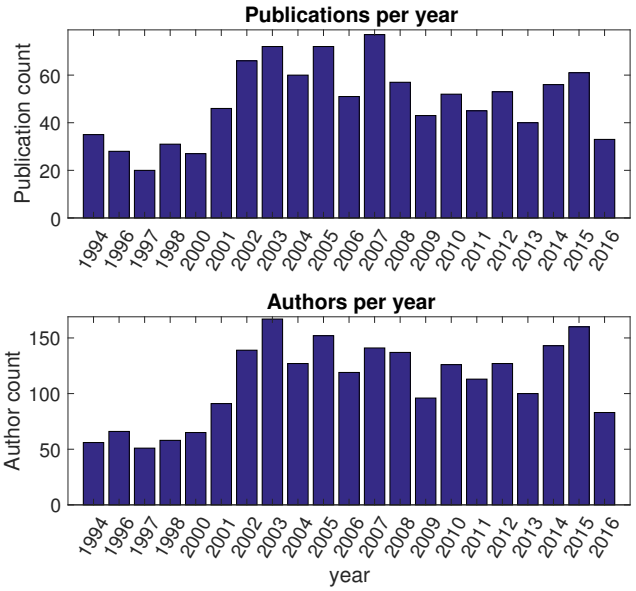


Figure 1: Evolution of the number of published ICAD papers per conference year (top). Evolution of the number unique authors per conference year (bottom)

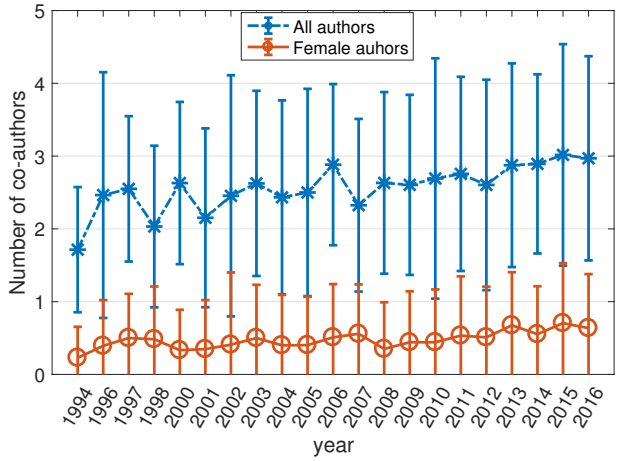


Figure 2: Evolution of the average number co-authors vs the average number of female authors per paper and conference year. Error bars mark the range of ± 1 Standard Deviation (std) from the mean.

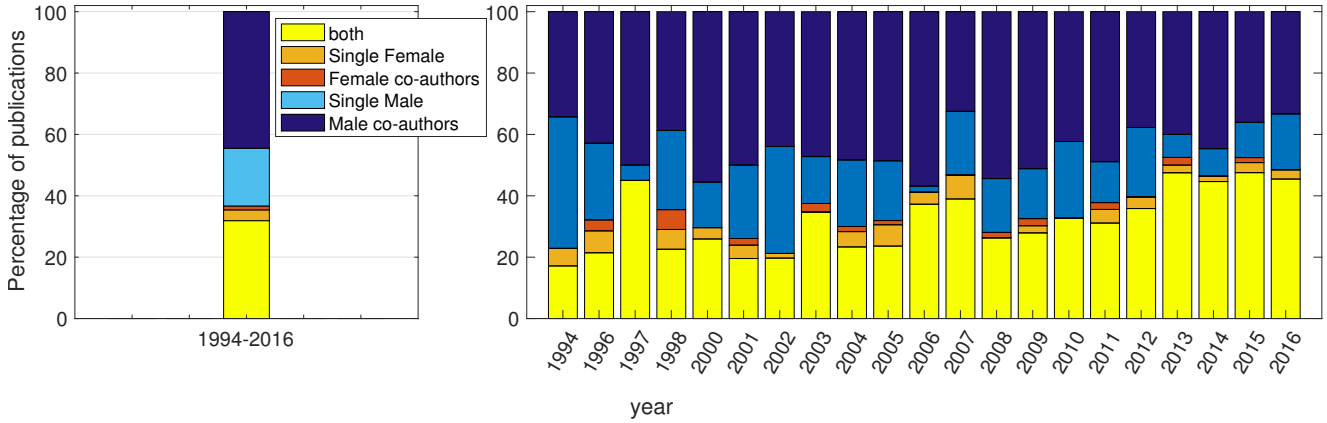


Figure 3: Percentage of publications with a) both male and female co-authors (yellow); b) single female author (light orange); c) only female co-authors (dark orange); d) single male author (light blue); e) only male co-authors (dark blue). Analysis is shown across all conference years (left) and per conference year (right)

authors either do not embrace this distinction or self-identify with the opposite gender to the one customary attributed to their first names. Inevitably, such cases cannot be accurately represented in the current data.

4. RESULTS

4.1. Female co-authorship in ICAD publications

Between 1994 and 2016 the International Community for Auditory Display has organized 21 conferences around the globe, 8 of which were held in the United States, 9 in Europe, 2 in Australia, 1 in Canada, and 1 in Asia. ICAD is a single track conference with an average of 48.8 papers (min = 20, max = 77, std = 16.16) and 110 authors (min = 51, max = 167, std = 36.4) per year. As can be seen in Figure 1, both the number of publications (top) and the corresponding number of unique authors (bottom) have the tendency to grow. This increase has been clearer since 2001, shortly after the conference started to travel globally.

Over the years, in addition to the growing number of publications, a shift towards more collaborative projects seems to have taken place. More specifically, as can be seen in Figure 2 while for the first conference in 1994 the average number of co-authors per publication was 1.7 (std = 0.86), by 2016 it reached 2.9 (std = 1.4), with a maximum of 3 (std = 1.5) in 2015. Nevertheless, this trend does not fully reflect the participation of female authors in ICAD publications. In 1994 the average number of female authors per paper was as low as 0.2 (std = 0.43). Even though this number has increased by a factor of 3.5 over the course of 21 conferences (Figure 2), reaching a maximum of 0.7 (std = 0.82) in 2015, it still remains at very low levels.

Figure 3 (left) shows the combined co-authorship analysis across all ICAD conferences since 1994. Overall, only 36.7% of all papers include at least 1 female co-author. In more detail, 63.3% of them were written exclusively by male collaborators, 31.9% were the result of a collaboration between male and female researchers, and only 4.8% were authored exclusively by female researchers. It is worth pointing out that 30% of the male authored papers were single-author publications, while for female authored publications that percentage was as high as 70%. In order to explore the evolution in the representation of female authors in ICAD publications, a co-authorship analysis per conference year was also performed. As can be seen in Figure 3 (right), with the exception of the 1997 and 2007 conferences, when the percentage of publications with at least one female author reached 45% and 46.8%, it was not until 2013 that more that half of the papers (52.5%) included female researchers.

In general, the number of published ICAD papers, co-authored by at least 1 female researcher, seems to be consistently increasing. The conference years with the lowest female representation were 1994 (22.9%) and 2002 (22.2%), and those with the highest were 2013 (52.5%) and 2015 (52.4%). The percentage of papers written exclussively by female authors does not seem to follow the same trend. The highest percentages of such papers, including single-author publications, appeared during the earlier ICAD years, prior to 2008, reaching a maximum of 13% in 1998. Since 2008, the corresponding percentages have varied between 1.5% and 6.6%. Another interesting observation is that the number of publications authored by a single female author is consistently larger than that of papers written by a group of female-only scientists.This observation implies, that female ICAD authors are more likely to publish as part of a mixed-gender author group or as single researchers, than forming female-exclusive groups. In contrast to that behavior, for male authors the formation of single-gender author-groups remains a very popular option throughout the ICAD conference years.

4.2. Female representation in ICAD publications

A more detailed analysis of female co-authorship in ICAD followed, looking at the percentages of papers with a) a leading female author, b) a last female author, c) both leading and last female authors, d) female co-authors and e) single authored papers. The choice to focus on leading and last authors is not coincidental, but rather driven by a customary academic convention, according to which first authors are usually scholars and last authors supervising researchers. Both roles are of particular interest, as they reflect the representation of prospective and established researchers within the ICAD community. We acknowledge that while this convention is well established within many scientific research do-

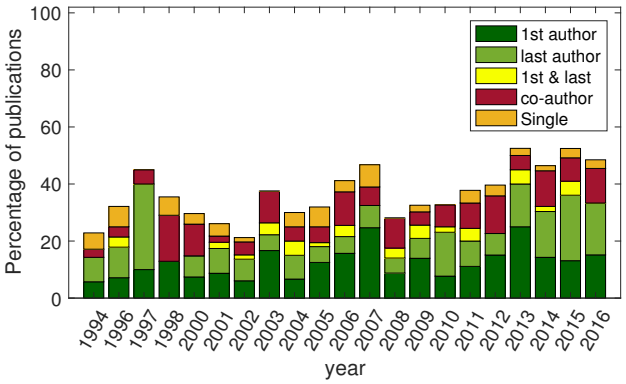


Figure 5: Analysis of publications including at least 1 female co-author, per conference year. In dark green: the percentage of papers with a leading female author; in light green: the percentage of papers with a last female author; in yellow: the percentage of papers with both leading and last female authors; in red: the percentage of papers with female co-authors in other roles; in orange: the percentage of single-author papers.

mains, it may not hold true for the humanities and the arts. Nevertheless, it still represents a portion of scholar work within the broad field of auditory displays, and is therefore worth exploring.

As can be seen in Figure 5, during the early conferences, the percentage of last female authors was considerably larger than that of leading authors. In the following years this relationship was occasionally inverted and approached an equilibrium in the last 5 conferences. Interestingly enough, throughout the history of ICAD conferences the percentage of publications with at least 1 female co-author was consistently smaller than the sum of percentages of papers with a leading and / or last female author. These findings imply that, even though in the recent years there exists an equivalent percentage of published ICAD papers with female scientists in leading or supervising roles, it is much less likely to encounter papers with female co-authors, for which the aforementioned research roles have been undertaken by male researchers. It is also worth pointing that the percentage of papers having both leading and last female authors was, with very few exceptions, smaller than any other of the aforementioned subgroups including single authored papers, fact which further supports the previously discussed observation that female researchers in the ICAD community are more likely to collaborate with other male colleagues than form female-exclusive research teams (see also Figure 3).

Up to this point, the data analysis focused on the number of published ICAD papers and the representation of female co-authors per publication and year. This perspective, even though informative as it reflects the growth and evolution of scientific collaborations within the ICAD community, does not really address the focal point of this paper which is the representation of women in ICAD conferences. An increase in the number of papers co-authored by female scientists might as well be the result of multiple concurrent collaborations of the same few female authors, rather than of an actual increase in the number of authors involved in the community. As a result, a different data analysis method was explored evaluating the percentage of female authors across all ICAD conferences.

Out of the 1520 unique authors who have published in ICAD conferences since 1994 only 20% were female (Figure 4-left). An analysis on their representation per conference year (Figure 4 (right)) revealed that levels have slightly increased but, in general, remain very low, ranging between 10.8% and 29% (mean = 18.7%, std = 4%). Such findings imply that the previously discussed increase in the participation of female authors in ICAD publications (Section 4.1) is related to the general growth of the annual conferences, and does not reflect an increase in the actual number of female scientist and artists who have chosen to publish with ICAD.

The 304 identified female authors who have published in ICAD conferences were affiliated with 188 institutions and / or companies from 25 countries worldwide. Table 1 shows the Top-10 countries ranked according to the number of female authors they have hosted. As can be seen, the vast majority of female authors (41.3%) have been associated with at least one institution in the USA at some point in their career. Other countries with relatively large female ICAD populations are: the UK, France, Canada, Austria, Germany, Australia, Finland, Japan, and Italy. Table 2 complements this information with a list of Institutions ranked according to the number of female ICAD authors they have hosted. Even though further discussion on this topic is not possible

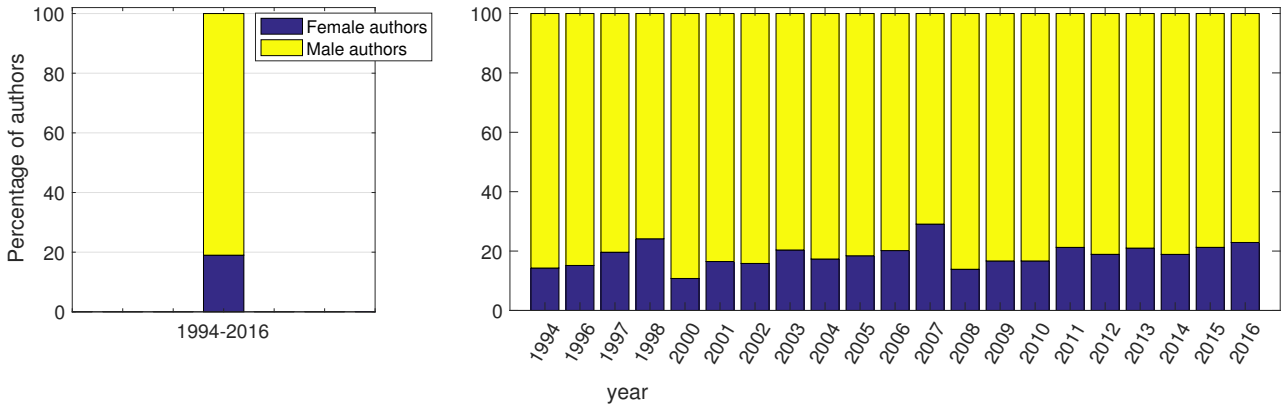


Figure 4: Percentage of unique female (blue) and male (yellow) authors, across all conference years (left), per conference year (right).

Table 1: Countries with the largest percepntage of female ICAD authors

Index	Country	% of Female Authors
1	USA	41.3
2	UK	14.8
3	France	7.4
4	Canada	5.5
5	Austria & Germany	4.2
7	Australia	3.5
8	Finland & Japan	3.2
10	Italy	2.9
11-25	Others	9.1

Table 2: Institution ranking according to the number of affiliated female ICAD authors.

Institution	# of Female Authors
Georgia Institute of Technology	9
University of York	9
CNRS	7
Michigan Technological University	7
University of British Columbia	7
University of Michigan	7
University of Glasgow	6
Boston University	5
IMASSA	5
Universita di Salerno	5

unless a similar analysis per country and institution is performed for all male ICAD authors, these two tables are included for future reference and discussion.

4.3. Returning ICAD community members

The very small fluctuations in the representation of male and female authors in ICAD conferences, discussed in Section 4.2, give rise to the following two questions: a) is the International Community for Auditory Display dominated by an increased number of returning researchers and artists who keep publishing their work in ICAD conferences , and b) does there exist a strong community of female researchers and artists who have chosen to follow and publish their work in ICAD conferences? In an attempt to address these points, distributions of recurrent male and female authors were created and compared using a two-sample unequal variance t-test. Results showed no significant difference between the two distributions ($t(553) = -1.16$, $p = 0.42$) at the default 5% significance level, and a Cohen's d effect size of $d = 0,07$, suggesting that both genders' "loyalty" to ICAD conferences is similar.

In order to assist visibility, the aforementioned distributions were converted into percentages and divided into the following 3 subgroups: percentage of male and female authors who published in the conference a) once, b) up to three times, and c) more than four times. As can be seen in Figure 6, 81% of all female and 77.1% of all male authors have published their work in a single ICAD conference, while 12.8.4% and 16.3% in up to 3, and 6.3% and 6.6% in 4 or more conferences, respectively. An investigation of potential reasons for the significantly large number of authors

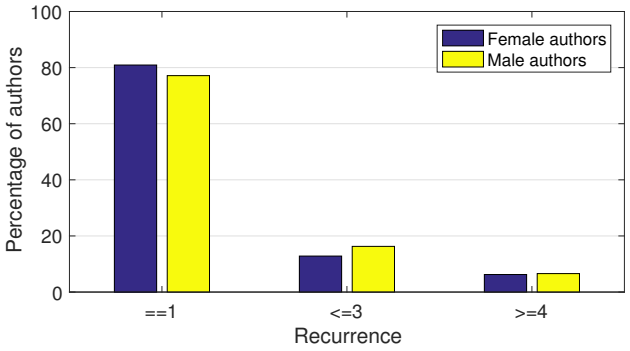


Figure 6: Comparison of the recurrence of female (blue) and male (yellow) authors in ICAD conferences over the years.

who have chosen to publish in a single ICAD conference, and how does it compare with that of other long-lasting conferences is beyond the scope of this paper, and should be revisited separately. Nevertheless, what remains of interest is the lack of evidence of any gender bias on ICAD "loyalty". Even though the absolute numbers of male and female published authors are significantly distinct, the choice of whether or not to follow this conference in its journey over the years does not seem to be affected by gender.

5. DISCUSSION AND FUTURE WORK

Over the past few years, there has been an increase in the number of publications studying women's involvement in music-production, composition, and audio-related research fields. To no surprise, albeit somewhat disappointingly, women's involvement in these artistic and scientific areas as been found to be alarmingly small. Several factors leading to this situation have been previously proposed, such as a) social expectations [6], b) lack of role-models [7], and c) issues related to work-life balance [8]. Most related literature concludes that, in order to encourage young female researchers and artists to be involved in music and audio-related fields, we need to study the achievements of experienced female researchers, and learn from their career paths. With this work, the authors wish to contribute to the understanding of women's representation in music and audio technology fields, by focusing on a study of female author participation in the conferences of the International Community for Auditory Display.

A simple gender analysis of all ICAD authors since 1994 revealed that women account for less than 20% of the published researchers across almost every conference. This percentage, even though low, is directly comparable to those of other related communities, such as the International Computer Music Association (ICMA) [2], and the International Society for Music Information Retrieval (ISMIR) [3] and almost double the percentage of female Audio Engineering Society (AES) members over the past decade, which is as low as 10%. The significant drop in the number of women affiliated with the AES community can possibly be attributed to its more technically oriented scope, which traditionally focused on audio signal processing and engineering.

While a more in-depth comparison between the studies of the aforementioned communities and the results of this paper cannot take place due to differences in data collection and analysis, their

statistics still serve as a very valuable comparison for the existing gender-related tendencies in the fields of music and audio research. The authors acknowledge that the study of female authors in ICAD conferences is not equivalent to a study of the representation of women in the International Community for Auditory Displays. Active involvement in a community can be manifested in various ways beyond scientific publications, such as conference attendance, conference organization, or even through the involvement with the community's Board of Directors. Future work will involve the collection and analysis of additional data related to the conference organization and attendance in an attempt to get a more complete view of women's involvement in the community. Additional points of interest include a study of the effect of geographic location on the infometrics of ICAD conferences, as well as the investigation of choice of research topics undertaken by male and female researchers.

6. REFERENCES

[1] C. Frauenberger and T. Stockman, "Auditory display desiginan investigation of a design pattern approach," *International Journal of Human-Computer Studies*, vol. 67, no. 11, pp. 907–922, 2009.

[2] M. Mathew, J. Grossman, and A. Andreopoulou, "Women in Audio: contributions and challenges in music technology and production," in *141st Audio Engineering Society Convention*, Los Angeles, USA, 2016, pp. 1–10.

[3] K. Choi, J. H. Lee, A. Laplante, Y. Hao, S. J. Cunningham, J. S. Downie, and U. Washington, "Wimir : an Informetric Study on Women Authors in Ismir," *Proc. 17th International Society for Music Information Retrieval Conference*, pp. 765–771, 2016.

[4] C. E. Seashore, "Why no great women composers?" *Music Educators Journal*, vol. 26, no. 5, pp. 21–88, 1940.

[5] A. McCartney, "Inventing images: constructing and contesting gender in thinking about electroacoustic music," *Leonardo Music Journal*, vol. 5, no. 1, pp. 57–66, 1995.

[6] ——, "In and out of the sound studio," *Organised Sound*, vol. 8, no. 01, pp. 89–96, 2003.

[7] K. De Welde, S. Laursen, and H. Thiry, "Women in science, technology, engineering and math (stem)," 2007.

[8] J. C. Williams, "The 5 biases pushing women out of stem," *Harvard Business Review*, 2015.

[9] C. R. Sugimoto, V. Lariviere, C. Ni, Y. Gingras, B. Cronin, *et al.*, "Global gender disparities in science," *Nature*, vol. 504, no. 7479, pp. 211–213, 2013.

[10] D. Maliniak, R. Powers, and B. F. Walter, "The gender citation gap in international relations," *International Organization*, vol. 67, no. 04, pp. 889–922, 2013.

[11] X. Hu, K. Choi, J. H. Lee, A. Laplante, Y. Hao, S. J. Cunningham, and J. S. Downie, "Wimir: An informetric study on women authors in ismir," in *Proceedings of the 17th International Conference on Music Information Retrieval (ISMIR)*, 2016.

A COMPARISON BETWEEN THE EFFICACY OF TASK-BASED VS. DATA-BASED sEMG SONIFICATION DESIGNS

Daniel Verona

Texas A&M University,
Department of Biomedical Engineering,
Emerging Technologies Building, 3120 TAMU,
101 Bizzell Street, College Station, TX 77843
daniel.j.verona@gmail.com

S. Camille Peres

Texas A&M University,
Department of Environmental and Occupational Health, School of Public Health
212 Adriance Lab Rd., College Station, TX 77843
peres@sph.tamu.edu

ABSTRACT

Historically, many sonification designs that have been used for data analysis purposes have been based on data characteristics and have not been explicitly based on the listener's task. These sonification designs have often been described as annoying, confusing, or fatiguing. In the absence of a generally accepted theoretical framework for sonification design, there is a need for improvements in sonification design as well as a need for empirical evaluation of task-based sonification designs. This research focuses on surface electromyography (sEMG) sonification and two sEMG data analysis tasks: determining which of two muscles contracts first and which of two muscles exhibits a higher exertion level. Both of these tasks were analyzed using a task analysis technique known as GOMS (Goals, Operators, Methods, Selection Rules) and two sonification designs were created based on the results of these task analyses. Two Data-based sEMG sonification designs were then taken from the sEMG sonification literature, and the four designs (2 Task-based and 2 Data-based) were empirically compared. Significant effects of sonification design on listener performance were found, with listeners scoring more accurately using the Task-based sonification designs. Based on these results, we argue for wider application of task analysis methods to sonification design and for the inclusion of task analysis methods into a generally accepted theoretical framework for sonification design.

1. INTRODUCTION

The auditory display community has known for quite some time that designing effective data sonifications is no small endeavor. A report published in 1999 highlighted the need for a sonification design method in order to establish a theoretical foundation upon which to base sonification designs [1]. Additionally, at the first ICAD conference in 1992, Sarah Bly referred to the lack of a theory of sonification as "a gaping hole impeding progress in the field" [2]. While strides have certainly been made since then towards establishing theories and guidelines for sonification design [3] [4] [5] [6] [7] [8] [9], a generally accepted sonification design framework, or theory of sonification design, still does not exist (S. Barrass, D. Brock, M. Gröhn, B. Walker, D. Worrall, personal communication, July 3, 2016).

https://doi.org/10.21785/icad2017.071

Arguably the most common type of sonification design in use today is parameter-mapping sonification, in which various parameters of sound (i.e. pitch, loudness, spatial location, timbre, etc.) are mapped onto data trends over a certain range and polarity. While this method of sonification can be effective for certain applications, it tends to result in sonifications that are annoying [10] [11] [12], confusing [13] [14], or fatiguing [15] [16]. Another problem with parameter-mapping sonification is that perceptual entanglement of various auditory parameters can lead to changes in one auditory dimension being perceived as changes in a different auditory dimension, which can obscure the meaning of a sonification [13]. This perceptual entanglement of various auditory parameters has led to what is known as The Mapping Problem, which is generally considered to be one of the primary obstacles currently facing sonification research [17].

Roddy and Furlong have proposed that improvements in sonification aesthetics could help to solve The Mapping Problem [18]. Sonification aesthetics have been understood in a variety of different ways over the years [9], and one important finding is that aesthetics and function cannot be treated independently in auditory display [10]. Building on this idea, Roddy and Furlong argue that sonification aesthetics deals primarily with meaning-making, and not with the overall "niceness" of the sound (which Roddy and Bridges refer to as sonification "cosmetics" [19]). Roddy and Furlong propose that looking to embodied cognition and creating sonifications that are mapped along embodied schemata may help listeners to derive the intended meaning from the sonification, and thus improve the aesthetic framework of the sonification. Embodied schemata are gestalt-like frameworks derived from the recurrent perceptual patterns encountered in daily life and they form the basic units of cognition. By mapping sonifications along these embodied dimensions, Roddy and Furlong argue that it could be possible to reconfigure entangled auditory dimensions into more comprehensible channels, thus circumventing The Mapping Problem and making sonifications that are more meaningful and easily understood [18]. The challenge they identified to this approach, however, is that the question of how to map specific embodied schemata to specific sonification tasks is not well understood.

Indeed, little research in general has been performed looking at how to design sonifications for specific tasks. This is due, in part, to the fact that many sonifications have not been explicitly designed to be tailored to the listener's task. Historically, many sonifications have been designed based on characteristics of the data being sonified (such as the type of data, number of data dimensions, number of data points etc.) [8], coupled with the designer's intuition [20] [21]. Additionally, many sonification designs to date have not been empirically evaluated [22], further compounding the

problematic lack of understanding regarding how to design sonifications for specific tasks.

While a number of researchers have discussed the importance of *considering* the task when designing a sonification [3] [5] [6] [23] [24] [25] [26], few have discussed in significant depth how to go about understanding the listener's task using task analysis techniques (with the notable exception of Barrass's dissertation [3]). Seemingly fewer still have created sonification designs that were based on a task analysis, and to date, we are not aware of any empirical evaluation of task-analysis-based sonification designs.

One recent study investigated the effects of sonification design on listener performance for two different sEMG data evaluation tasks [27]. Results showed that one sonification design (Pitch/Loudness) yielded the most accurate listener performance for the first task (identifying which of two muscles contracted first), while a different design (Pitch/Loudness/Attack time) yielded the most accurate listener performance for the second task (identifying which of two muscles exhibited a higher exertion level).

Given this finding that sonification design efficacy was task-dependent, it seemed a natural progression to begin investigating task-specific sonification designs. This was done by using task analysis methods as a means for establishing sonification design criteria.

This paper presents the results of this investigation into task-based sonification designs which involved:

- Identifying an appropriate task analysis technique that would provide the needed information
- Performing task analyses of the two sEMG data evaluation tasks used in the study
- Creating two sonification designs based on the results of these task analyses
- Comparing these Task-based designs to two Data-based designs taken from the sEMG sonification literature

2. TASK ANALYSIS AND GOMS

A task analysis is a design tool commonly used in fields such as Human Factors and Human Computer Interaction [28] [29]. Its purpose is to provide knowledge about users, their goals in accomplishing a task, their environment, the manual elements of the task, the cognitive elements of the task, the tools used to perform the task, the duration, order, and complexity of the task, as well as any other unique factors pertaining to the task [30].

The idea of using task analysis techniques to inform sonification design is not new, as previously mentioned. Barrass discussed the use of task analysis methods at length as part of his TaDa framework for sonification design [3]. Anderson also proposed the incorporation of task analysis methods into a framework for sonification design [5]. Walker and Nees stated that effective sonifications will require an understanding of the listener's function and goals within a system, that the task is a crucial consideration for the success or failure of a sonification, and that a display designer's knowledge of the task will constrain and inform the design of a sonification [6].

One issue we have not seen addressed in the sonification literature, however, is a discussion regarding what type of task analysis technique to use for informing sonification design. Many types of task analysis techniques have been developed, and each technique provides the designer with somewhat different information. One broad way to categorize the many different types of task analysis methods is to divide them into action-oriented methods and cognitive methods [31]. Action-oriented methods (such as the commonly used hierarchical task analysis, or HTA) focus on observable actions, or identifying, in top down fashion, the goal of the task, as well as the various subtasks and conditions under which those subtasks must be performed in order to achieve the goal. Cognitive methods, on the other hand, focus on analyzing and outlining the unseen mental processes – diagnosis, decision making, problem solving, etc. – that can give rise to human error [31].

Based on this categorization, it at first seemed reasonable to take a cognitive approach to task analysis for use in informing sonification design, since comprehension of a sonification does not depend on observable actions, but rather on unseen mental operations. However, formal cognitive task analysis (CTA) methods may not be feasible for use in sonification design due to the fact that they typically require observation of expert performance, interviews with subject-matter-experts (SME's), or capturing an expert's performance with a think aloud protocol or subsequent recall [32]. Since interpreting a sonification is a primarily cognitive task, it would not be possible to visually observe an expert's performance of a sonification interpretation task to gain much useful information. To compound the problem, doing a think aloud protocol in real time for the interpretation of an EMG sonification would be difficult considering the interference speaking would have on listening to the sonifications.

To account for the cognitive aspects inherent to sonification interpretation, and avoid the ways in which CTA methods may not be ideal for decomposing sonification interpretation tasks, we are interested in the use of GOMS for informing sonification design. GOMS stands for Goals, Operators, Methods, and Selection Rules, and it is a form of HTA originally developed by Card, Moran, and Newell [33]. Its aim is to model and predict user behavior, or in the case of sonification, listener behavior. The four components of a GOMS model are as follows [34]:

- **Goals**: what the user is trying to accomplish. Goals can be, and often are, decomposed into Goal/Sub-goal hierarchies.
- **Operators**: actions performed in service of a goal. Operators can be perceptual, cognitive, or motor acts, or some combination of these.
- **Methods**: sequences of operators and sub-goal invocations that accomplish a goal.
- **Selection Rules:** when there is more than one method for accomplishing a goal, selection rules are the rules that the user employs to determine which method to use to accomplish the goal.

3. APPLICATION OF GOMS TO IDENTIFY SONIFICATION DESIGN CRITERIA

In this study, participants were asked to listen to sonifications of two channels of sEMG data, referred to as Muscle A and Muscle B, respectively. In the sonifications, both Muscle A and Muscle B began at rest, contracted at close to the same time, remained contracted for a few seconds, and then returned to rest. After listening to each sonification, participants were asked to perform the following two sEMG data evaluation tasks:

Goal: DETERMINE IF A OR B CONTRACTS FIRST, OR IF THEY CONTRACT SIMULTANEOUSLY

Method for TIME Goal:

SG 1. Start Task
SG 2. Identify 1st Muscle Activation
SG 3. Determine if 1st Activation was Muscle A or Muscle B
SG 4. Determine if other Muscle Activated also
SG 5. If Unsure regarding Subgoal 3, Identify 2nd Muscle Activation
SG 6. Determine if 2nd Activation was A or B
SG 7. Determine if A or B Contracted First
SG 8. Report if A or B Activated First

Method for Subgoal 1:
Start Task

Op 1. Grasp computer mouse
Op 2. Point with mouse to PLAY button
Op 3. Left-click PLAY button

Method for Subgoal 2:
Identify 1st Muscle Activation

Op 1. Perceive sonic event indicating muscle activation
Op 2. Place sonic event in auditory store
Op 3. Shift attention to auditory store

Method for Subgoal 3:
Determine if 1st Activation was A or B

Op 1. Perceive unique sonic identifier for A or B
Op 2. Equate sonic identifier with A or B
Op 3. Place identification of A or B into working memory

Method for Subgoal 4:
Determine if other Muscle Activated also

Op 1. Sonic event indicating other muscle activating simultaneously perceived?
Op 2. If yes, store this knowledge in working memory
Op 3. If no, then keep identification of A or B (from Subgoal 3) in working memory

Method for Subgoal 5:
If Unsure regarding Subgoal 3, Identify 2nd Muscle Activation

Same as Method for Subgoal 2, but for second muscle activation

Method for Subgoal 6:
Determine if 2nd Activation was A or B

Same as Method for Subgoal 3, but for the second muscle activation

Method for Subgoal 7:
Determine if A or B Contracted First

Op 1. Retrieve identification of first contraction as Muscle A or B from working memory (Subgoal 3)
Op 2. If second muscle contraction was perceived simultaneously, retrieve this knowledge from working memory (Subgoal 4)

Method for Subgoal 8:
Report if A or B Contracted First

Op 1. Grasp computer mouse
Op 2. Point with mouse to radio button indicating correct answer
Op 3. Left-click radio button

Goal: DETERMINE IF A OR B HAS A HIGHER EXERTION LEVEL, OR IF THEY ARE THE SAME

Method for LEVEL Goal:

SG 1. Start Task
SG 2. Identify Muscle A's Activation
SG 3. Identify Muscle B's Activation
SG 4. Monitor A's Exertion Relative to B's Exertion during muscle contraction
SG 5. Identify when A Returns to Rest
SG 6. Identify when B Returns to Rest
SG 7. Determine if A or B had Higher Exertion Level
SG 8. Report if A or B had Higher Exertion Level

Method for Subgoal 1:
Start Task

Op 1. Grasp computer mouse
Op 2. Point with mouse to PLAY button
Op 3. Left-click PLAY button

Method for Subgoal 2:
Identify Muscle A's Activation

Op 1. Perceive sonic event
Op 2. Perceive unique sonic identifier for Muscle A
Op 3. Place sonic event in auditory store
Op 4. Shift attention to auditory store
Op 5. Equate identifier with Muscle A

Method for Subgoal 3:
Identify Muscle B's Activation

Same as Method for Subgoal 2, but for Muscle B

Method for Subgoal 4:
Monitor A's Exertion Relative to B's Exertion

Op 1. Use echoic memory to continuously update A's max exertion
Op 2. Use echoic memory to continuously update B's max exertion
Op 3. Place max exertion in working memory

Method for Subgoal 5:
Identify when A Returns to Rest

Op 1. Perceive sonic event indicating Muscle A returning to rest
Op 2. Place sonic event in auditory store
Op 3. Shift attention to auditory store
Op 4. Stop continuously updating max exertion for Muscle A

Method for Subgoal 6:
Identify when B Returns to Rest

Same as Method for Subgoal 5, but for Muscle B

Method for Subgoal 7:
Determine if A or B had a Higher Exertion Level

Op 1. Retrieve max exertion level from working memory
Op 2. Equate max exertion level with Muscle A or Muscle B

Method for Subgoal 8:
Report if A or B had a Higher Exertion Level

Op 1. Grasp computer mouse
Op 2. Point with mouse to radio button indicating correct answer
Op 3. Left-click radio button

Figure 1 – The left column shows the GOMS analysis for the TIME task (determining which of two muscles contracts first), and the right column shows the GOMS analysis for the LEVEL task (determining which of two muscles exhibits a higher exertion level). In these graphs, "SG" stands for 'Subgoal' and "Op" stands for 'Operator'.

1. **TIME Task**: Identify which muscle (A or B) contracted first
2. **LEVEL Task**: Identify which muscle (A or B) exhibited a higher exertion level

To design sonifications specifically for these two tasks, GOMS analyses were performed for both tasks, and the results are shown above in Figure 1. These GOMS analyses only show Goals, Subgoals, and Operators. The Method is to follow the Subgoals in numerical order, and for each Subgoal to follow the Operators in numerical order. The assumption is that there are not additional Methods that would allow for the accomplishment of each Goal, and thus there are no Selection Rules shown for selecting between competing Methods. Identification of the various Subgoals involved for each task served as the primary factor in establishing sonification design criteria for each task.

3.1. Design criteria for the TIME task

For the task of identifying which muscle contracts first, the analysis in the left column of Figure 1 indicates that a listener must be able to understand that the task has started (Subgoal 1), identify when the first muscle changes state from rest to contraction (Subgoal 2), then determine if that muscle was Muscle A or Muscle B (Subgoal 3). If the sonification does not give the listener the ability to accomplish even one of these Subgoals, the listener will not be able to complete the task. Thus, the design criteria for the sonifications based on the GOMS model for the TIME task are:

1. The start of the listening task must be evident
2. The sound of the first muscle changing state from rest to contraction must be evident
3. The listener must have a way of distinguishing between the sound of Muscle A activating and the sound of Muscle B activating

3.2. Design criteria for the LEVEL task

For the task of identifying which muscle has a higher exertion level, the analysis shown in the right column of Figure 1 indicates that the listener must be able to understand that the task has started (Subgoal 1), determine when both muscles change state from rest to contraction (Subgoals 2 and 3), monitor the exertion level difference between Muscle A and B for the duration of their contractions (Subgoal 4), identify when both muscles revert back to rest (Subgoals 5 and 6), then determine if Muscle A or B had a higher exertion level (Subgoal 7). Once again, failure to accomplish any of these Subgoals will prevent the listener from completing the task. Thus, the design criteria for the sonifications based on the GOMS model for the LEVEL task are:

1. The start of the listening task must be evident
2. The sound of both muscles changing state from rest to contraction must be evident
3. The exertion level difference between the two muscles must be evident
4. The sound of both muscles changing state from contraction back to rest must be evident

4. METHODS

Each sonification design created for this study was coded in the SuperCollider audio synthesis environment. All sEMG data processing (rectifying and filtering) was performed using MATLAB.

4.1. Study design

This study compared the efficacy of two Task-based sonification designs to two Data-based sonification designs taken from the EMG sonification literature, for two different tasks – muscle activation time and muscle exertion level. There were thus three main independent variables (IVs: Data-based design, Task-based design, and Task) with two levels of each variable. Further, there were 4 levels of difficulty for each task, adding another IV. For the Data-based designs, the first level was a pitch mapping and the second level was a loudness/timbre mapping, referred to henceforth as a loudness mapping. These designs were taken from a 2012 study investigating sonification of EMG data for use in analyzing human movements [35]. The details of these designs are explained below in section 4.2. For the Task-based designs, the first level was the "Task-Panning" design which used short beeps to indicate the onset of muscle activation and a panning tone to indicate exertion level difference. The second level was the "Task-Filter" design which also used short beeps to indicate the onset of muscle activation, but used a panned filter cutoff mapping to indicate muscle exertion difference. There were two dependent variables associated with the two levels of the task IV to assess performance for each design: judgment of muscle activation time (TIME task) and judgment of muscle exertion level (LEVEL task). The IV's and Levels are described in Table 1 below.

Table 1: IV and Level for the four sonification designs and two tasks.

IV 1: Data-based	IV 2: Task-based	IV 3: Task
Data-Pitch	Task-Panning	Muscle activation time difference
Data-Loudness	Task-Filter	Muscle exertion level difference

This study was a within-subjects factorial design. Participants listened to 16 sonifications with each of the four designs for a total of 64 sonifications. The presentation order of the four sonification designs was counterbalanced to account for training effects.

4.2. Data-based designs

As previously mentioned, the two Data-based designs used in this study were taken from a 2012 paper by Matsubara et al. [35]. We chose to use designs from Matsubara's paper because participants in Matsubara's study were asked to perform sEMG data evaluation tasks that were similar to the sEMG data evaluation tasks that we asked our participants to perform. There were three design methods used in Matsubara's study: Method A: Pitch, Method B: Loudness/Polyphonic Timbre, and Method C: Timbre. Methods A and B were chosen as the Data-based designs for

this study because they resulted in the best listener performance during Matsubara's study.

The Data-Pitch design was created according to the specifications laid out in Matsubara [35] for Method A, with the first channel of sEMG data (Muscle A) sonified using a sine wave tone over a frequency range of 300-525 Hz, and the second channel of sEMG data (Muscle B) sonified using a sine wave tone over a lower frequency range of 165-345 Hz. Additionally, we decided to spatialize this design by panning the first channel of sEMG data (A) hard left and panning the second channel of sEMG data (B) hard right. We made this decision based on our previous findings that spatialization helps listeners distinguish between sEMG channels [27].

The Data-Loudness design was also created according to the specifications laid out in Matsubara's paper for Method B. Again, we spatialized the design in an attempt to enhance listener performance in keeping with our previous findings.

4.3. Task-based designs

The Task-Panning design was based on the design criteria for the TIME and LEVEL tasks from the task analyses. To ensure that the listener would know that the sonification was playing, a soft, low-pass-filtered white noise was played while the muscles were at rest. The cutoff frequency of the LPF was set to 1000 Hz.

To indicate when each muscle activated, short beep tones were played when each muscle began to contract. To indicate the contraction of Muscle A, a short beep (0.07 sec duration) using a triangle wave at a frequency of 440 Hz was played in the left ear. To indicate the contraction of Muscle B, a short beep of equal duration using a triangle wave at a frequency of 330 Hz was played in the right ear. Once both muscles had begun to contract, the LPF white noise was turned off and a tone indicating exertion level difference began to play.

To indicate the exertion level difference between Muscle A and B, the sonification code calculated the difference in amplitude between A and B (Amp_A - Amp_B), and then mapped this difference to the pan position of a tone that played during muscle contraction. If the difference was positive, this meant that Muscle A had a higher exertion level and the tone panned left, and vice versa. When the difference in exertion was small (~0.05 V), the tone panned slightly left or right (to a value of +/– 0.7 on SuperCollider's Pan2 function). When the difference in exertion was larger (> 0.1 V), the tone panned hard left or right.

After muscle contraction, the tone became silent and the white noise returned to indicate that the muscles had returned to rest.

The Task-Filter design was also based on the design criteria for the TIME and LEVEL tasks from the task analyses. For this design, when the muscles were at rest, a soft, low-pass-filtered sawtooth wave was played, one in the left channel to represent Muscle A and one in the right channel to represent Muscle B. The frequency of the waves was 100 Hz, and the cutoff frequency of the LPF was 300 Hz. The two waves were played at equal amplitude so as to be perceived in the center of the stereo field.

To indicate when each muscle activated, short beep tones were played right when each muscle began to contract. To indicate the contraction of Muscle A, a short beep (0.09 sec duration) using an additive synthesis tone with a fundamental frequency of 300 Hz was played in the left ear. To indicate the contraction of Muscle B, the same short beep was played in the right ear. The fundamental frequency of 300 Hz was chosen so that these beeps would "sit on top of" the sawtooth wave tones (which were LPF'd at 300 Hz) and not interfere with them.

To indicate the exertion level difference between Muscle A and B, the sonification code calculated the amplitude difference in the same manner as in the Task-Panning design. If the difference was positive, this meant that Muscle A (in the left channel) had a higher exertion level and the difference was mapped to the cutoff frequency of the LPF in the left channel, such that the cutoff increased to allow more high frequency content to be heard in the left channel during muscle contraction. The opposite occurred when the amplitude difference was negative, with the cutoff of the right channel's LPF increasing to indicate that Muscle B had a higher exertion level. For small exertion differences (0.05 V), the cutoff would increase from 300 Hz to 1200 Hz, and for larger exertion differences (0.15 V), the cutoff increased from 300 Hz to 3600 Hz.

After muscle contraction, the cutoff of both LPF's was set back to 300 Hz to indicate that the muscles had returned to rest.

4.4. Activation time/Exertion level differences

For each of the four sonification designs, participants listened to 16 sonifications. Of these 16, 4 displayed both muscles activating at the same time, and 4 each displayed both muscles activating 0.13 sec apart, 0.26 sec apart, and 0.39 sec apart.

Additionally, out of the 16, 4 sonifications displayed both muscles exhibiting the same exertion level, and 4 each displayed both muscles exhibiting a 0.05 V, 0.10 V, and 0.15 V amplitude difference during muscle contraction. The 16 sonifications for each design were numbered according to Table 2 below.

Table 2. Listing of structure for the 16 sonifications for each design.

		Activation time difference			
		0 sec	0.13 sec	0.26 sec	0.39 sec
Exertion level difference	0 V	*1*	*2*	*3*	*4*
	0.05 V	*5*	*6*	*7*	*8*
	0.10 V	*9*	*10*	*11*	*12*
	0.15 V	*13*	*14*	*15*	*16*

As an example, Sonification #1 for any given design displayed both muscles contracting at the same time (0 sec activation time difference) and exhibiting the same exertion level (0 V amplitude difference during contraction). Similarly, Sonification #11 in any given design displayed a 0.26 sec difference between the activation of Muscle A and the activation of Muscle B, and a 0.1 V difference in amplitude between Muscle A and Muscle B. The order in which each sonification within a given design was presented was randomized for each counterbalance.

4.5. Participants

Forty students and faculty from Texas A&M university participated in this study (27 male, 16 female, ages 19-59). They all self-reported as not having any hearing impairment that would interfere with their ability to participate. At the beginning of each session, participants signed a consent form, completed a demographic survey, and were asked

about their knowledge of and experience with EMG data. After this, they were briefly trained on what sEMG data is, what sonification is, and how sEMG data can be sonified.

4.6. Computer/Audio setup

The study was run locally through a browser (Google Chrome) using the XAMPP environment in conjunction with a MySQL database for recording participant responses. Participants listened to the sonifications through a pair of Beyerdynamic DT 770 Pro headphones.

4.7. Measures

Listener accuracy was measured as a proportion of correct responses for both tasks. After listening to each sonification, participants were asked two multiple choice questions, one each for the TIME and LEVEL tasks. The choices were:

1. Muscle A activated first (or had a higher exertion)
2. Muscle B activated first (or had a higher exertion)
3. A and B had the same activation time (or exertion level)
4. Unsure

For example, if a listener correctly identified if Muscle A or B contracted first for 8 out of the 16 Data-Pitch sonifications, their score was 8/16 = 0.5 for that Design/Task pair.

5. RESULTS

5.1. Overall performance

As seen in Figure 2, there was no effect of Task, $F(1, 42) = 1.782$, $p = 0.189$. However, there was a main effect of design, $F(2.079, 87.29) = 91.23$, $p < 0.001$, eta squared 0.69, and an interaction between Task and Design $F(2.55, 107.23) = 32.83$, $p < 0.001$, eta squared = 0.44.

Bonferroni pairwise comparisons indicated that performance was different based on design with the Data-Pitch design having the worst performance and Task-Filter having the best ($p < 0.001$). Data-Loudness and Task-Panning had performance levels in between those two and Bonferroni pairwise comparisons show that performance on all designs were significantly different from each other (all p's < 0.001). As shown in Figure 1, there was an interaction between Design and Task with the Data-Pitch design having better performance for the TIME task ($p < 0.034$), and the Task-Filter design having better performance for the LEVEL task ($p < 0.028$).

5.2. Performance by difficulty level

Figure 3 shows the results by difficulty level for the activation time task. This figure shows that performance differed by Design with the Task-based designs resulting in better performance than the Data-based designs (all p's < 0.01). The Task-based designs and Data-based designs were not different from each other (p's > 0.29), $F(2.246, 94.318) = 19.60$, $p < 0.001$, eta squared = 0.318. Figure 2 also shows that there were overall differences in performance based on the Activation Time Differences (ATD) with better performance when the differences

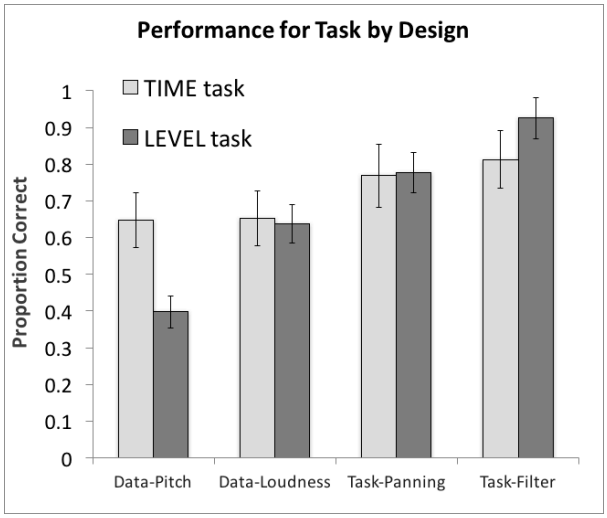


Figure 2 – Overall listener performance for each Design and for both Tasks

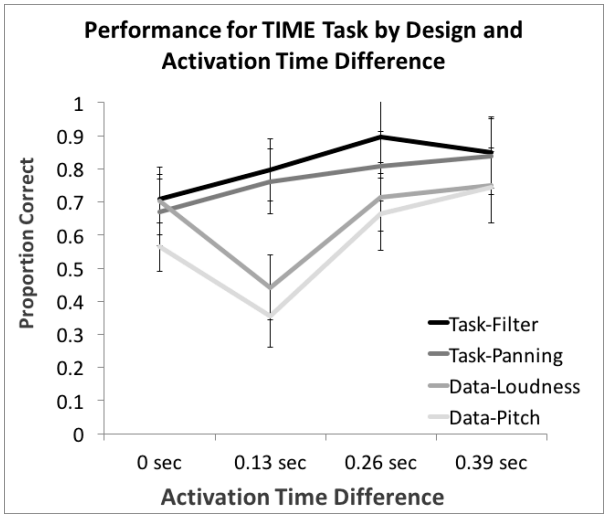


Figure 3 – Listener performance for the TIME Task for each Design and Activation Time Difference (ATD). ATD = time difference between activation of Muscle A and Muscle B

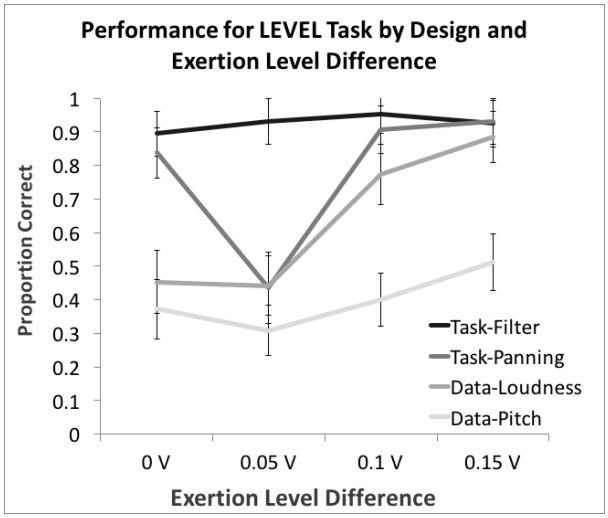


Figure 4 – Listener performance for the LEVEL task for each Design and Exertion Level Difference (ELD). ELD = amplitude difference during contraction between Muscle A and Muscle B

were larger (0.26 sec and 0.39 sec) (all p's < 0.001), F(1.539, 64.65) = 12.27, p < 0.001, eta squared = 0.23. The differences by Level differed by Design for the TIME task with Bonferroni comparisons indicating that Data-Pitch (0.13 sec) was different than all others and Data-Pitch (0 sec) was different than Data-Pitch (0.39 sec); Data-Loudness (0.13 sec) was different than all others; Task-Panning showed no performance differences by level; and Task-Filter (0.13 sec) was different than Task-Filter (0.26 sec), F(5.43, 228.11) = 7.68, p < 0.001, eta squared = 0.12.

Figure 4 shows that overall performance for the LEVEL task differed by Design with the Task-Filter Design (all p's < 0.01) resulting in the best performance and Data-Pitch design resulting in the worst. Bonferroni comparisons showed that all designs were different from each other with performance on the Task-Panning being lower than Task-Filter and greater than Data-Loudness, F(2.37, 99.55) = 154.54, p < 0.001, eta squared = 0.79. Figure 4 also shows that there were differences in performance based on the Exertion Level Differences with performance generally increasing as exertion level differences increased (all p's < 0.038). The exception to this was the decrease in performance from 0 V difference to 0.05 V difference, F(1.95, 81.91) = 38.12, p < 0.001, eta squared = 0.476. The differences by Level differed by Design with Bonferroni comparisons indicating that Data-Pitch (0.05 V) was different than Data-Pitch(0.15 V), Data-Loudness (0.15 V) was different than Data-Loudness (0, 0.05, 0.1 V), Data-Loudness (0, 0.05 V) was different than Data-Loudness (0.1, 0.15 V), Task-Panning (0.05 V) was different than Task-Panning (0, 0.1, 0.15 V), and Task-Filter showed no differences between levels, F(5.563, 233.648) = 15.534, p < 0.001, eta squared = 0.27.

6. DISCUSSION

The results of this study clearly indicate that for interpreting sEMG sonifications for these tasks, using sonification designs based on the task results in superior performance, particularly for the TIME task.

For the TIME task, The Data-Pitch and Data-Loudness designs showed poor performance when the activation time difference was 0.13 sec. By contrast, the Task-Panning and Task-Filter designs showed performance that essentially increased as the TIME difference increased (Figure 3). This was likely due to the fact that the Task-based designs were designed specifically to create a large, temporally precise contrast between the sound of a muscle at rest and the sound of a muscle beginning to contract. The Data-based designs did not provide the same level of perceivable contrast between the sound of a muscle changing state from rest to activation.

For the LEVEL task, there were interesting interactions based on the difficulty level of task with the more difficult stimuli (.05 V) reducing performance remarkably more with the Task-Panning design than any of other of the designs. Further, there were differences in performance between the Data-Pitch and Data-Loudness designs for the LEVEL task. This is likely due to two things: the Data-Pitch design used different pitch ranges for Muscles A and B which made a direct comparison between the two difficult, and the Data-Loudness design essentially made use of a panning effect by mapping muscle exertion level to loudness. Since the designs were spatialized into left and right audio channels, at larger exertion level differences (0.1 and 0.15 V), the Data-Loudness design acted like a panning mapping, and indeed, the Data-Loudness design showed similar performance for the LEVEL task as both of the Task-based designs, which both made explicit use of panning (see Figure 4).

These findings that Task-based designs can result in better listener performance than Data-based designs strongly suggest the broader integration of task-based approaches into the sonification design problem space. Additionally, they indicate that the inclusion of task analyses within a theoretical framework for sonification design may facilitate the development of this illusive framework.

Task-based approaches to sonification design do not seem to be well represented in the auditory display literature. It is not uncommon in the EMG sonification literature, for example, to see an explanation for *how* a sonification was designed but to not see an explanation for *why* it was designed that way. Justifications for design decisions are sometimes given, but they rarely seem to go beyond appeals to sonic cosmetics or "traditional" mappings like pitch and loudness.

A task-based approach to sonification design could allow sonification designers to use Human Factors and HCI design methodology to identify sonification design criteria. In so doing, this approach could afford sonification designers stronger justification for design decisions, as well as facilitate easier communication between sonification designers and HF/HCI researchers – which could broaden the ICAD community's impact and stimulate wider interest in the field.
As mentioned in the introduction to this paper, Roddy and Furlong have discussed sonification aesthetics and the problem of a disembodied approach to sonification design. They have argued that leveraging knowledge of embodied cognition and embodied schemata may help sonification designers to circumvent the Mapping Problem by mapping sonifications along embodied dimensions. Task-based approaches to sonification design may not be *embodied* in and of themselves, but since task analyses can provide in-depth knowledge of a user's task, and since mapping sonifications along embodied dimensions requires a deep understanding of the user's task, it seems that task-based approaches to sonification design may aid in identifying useful embodied schemata along which to map sonifications for specific tasks.

In conclusion, task analysis techniques are well established in fields such as Human Factors and HCI, where design decisions are critical. In this study, implementing task analysis techniques into the design of auditory displays was shown to be an effective approach for creating interpretable sonifications. Further use of task analysis techniques in auditory display is thus recommended. This study has served as a "proof-of-concept," and we believe that further use of task-based approaches in sonification research may help to ultimately ground sonifications in a more accessible – and perhaps embodied – aesthetic framework, thus leading to the development of more easily interpretable sonifications. If this is done, it could broaden the ICAD community's impact and generate wider awareness of, and interest in, the field.

7. AKNOWLEDGMENT

We wish to thank Dr. Masaki Matsubara at the University of Tsukuba for use of his sonification designs and MATLAB code.

8. REFERENCES

[1] Kramer, B. Walker, T. Bonebright, P. Cook, J.

Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, G. Evreinov, W. Fitch, M. Gröhn, S. Handel, H. Kaper, H. Levkowitz, S. Lodha, B. Shinn-Cunningham, M. Simoni, and S. Tipei, "Sonification Report: Status of the Field and Research Agenda," 1999.

[2] S. Bly, "Multivariate Data Mappings," in *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer, Ed. Westview Press, 1994, pp. 405–416.

[3] S. Barrass, "Auditory Information Design," *Aust. Natl. Univ. Aust.*, p. 288, 1997.

[4] B. Walker, "Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification," *J. Exp. Psychol. Appl.*, vol. 8, no. 4, pp. 211–221, 2002.

[5] J. Anderson, "Creating an Empirical Framework for Sonification Design," in *Proceedings of ICAD 05 - Eleventh Meeting of the International Conference on Auditory Display*, 2005, pp. 393–397.

[6] B. Walker and M. Nees, "Theory of Sonification," in *Principles of Sonification: An Introduction to Auditory Display and Sonification*, 2011, pp. 1–32.

[7] S. Barrass, "Sonification Design Patterns," *Int. Conf. Audit. Disp. 2003*, no. July, pp. 170–175, 2003.

[8] A. de Campo, "A Data Sonification Design Sace Map," in *Proceedings of the 2nd International Workshop on Interactive Sonification*, 2007, pp. 1–4.

[9] S. Barrass and P. Vickers, "Sonification Design and Aesthetics," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. Neuhoff, Eds. 2011, pp. 145–171.

[10] G. Leplatre and I. McGregor, "How to Tackle Auditory Interface Aesthetics? Discussion and Case Study," in *Proceedings of ICAD 04 - Tenth Meeting of the International Conference on Auditory Display*, 2004.

[11] C. Henkelmann, "Improving the aesthetic quality of realtime motion data sonification," *Comput. Graph. Tech. Rep.*, pp. 1–133, 2007.

[12] G. Baier, T. Hermann, and U. Stephani, "Event-based sonification of EEG rhythms in real time," *Clin. Neurophysiol.*, vol. 118, no. 6, pp. 1377–1386, 2007.

[13] D. Worrall, "Parameter Mapping Sonic Articulation and the Perceiving Body," in *The 16nth International Conference on Auditory Display*, 2010, pp. 207–214.

[14] S. C. Peres, "A Comparison of Sound Dimensions for Auditory Graphs: Pitch Is Not So Perfect," *J. Audio Eng. Soc.*, no. July, pp. 561–567, 2012.

[15] P. Vickers, C. Laing, M. Debashi, and T. Fairfax, "Sonification Aesthetics and Listening for Network Situational Awareness," *SoniHED - Conf. Sonification Heal. Environ. Data*, 2014.

[16] S. Pauletto and A. Hunt, "The Sonification of EMG Data," *Proc. 12th Int. Conf. Audit. Disp.*, pp. 152–157, 2006.

[17] D. Worrall, "A Method for Developing an Improved Mapping Model for Data Sonification," in *The 17th International Conference on Auditory Display*, 2011.

[18] S. Roddy and D. Furlong, "Embodied Aesthetics in Auditory Display," *Organised Sound*, vol. 19, no. Special Issue 01. pp. 70–77, 2014.

[19] S. Roddy and B. Bridges, "Sounding Human with Data: The Role of Embodied Conceptual Metaphors and Aesthetics in Representing and Exploring Data Sets," pp. 64–76, 2016.

[20] M. Nees and B. Walker, "Listener, Task, and Auditory Graph: Toward a Conceptual Model of Auditory Graph Comprehension," in *Proceedings of the 13th International Conference on Auditory Display*, 2007, pp. 266–273.

[21] B. Walker and G. Kramer, "Sonification Design and Metaphors: Comments on Walker and Kramer, ICAD 1996," *ACM Trans. Appl. Percept.*, vol. 2, no. 4, pp. 413–417, 2005.

[22] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," *PLoS One*, vol. 8, no. 12, 2013.

[23] G. Kramer, *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Westview Press, 1994.

[24] J. Flowers, "Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions," in *Proceedings of ICAD05 - Eleventh Meeting of the International Conference on Auditory Display*, 2005, pp. 406–409.

[25] P. Sanderson, J. Anderson, and M. Watson, "Extending Ecological Interface Design to Auditory Displays," in *Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia*, 2000, pp. 259–266.

[26] B. N. Walker, J. Lindsay, and J. Godfrey, "The Audio Abacus: Representing Numerical Values with Nonspeech Sound for the Visually Impaired," in *ASSETS*, 2004.

[27] S. C. Peres, D. Verona, T. Nisar, and P. Ritchey, "Towards a Systematic Approach to Real-Time Sonification Design for Surface Electromyography," *Displays*, vol. 47, pp. 25–31, 2017.

[28] D. Phipps, G. Meakin, P. Beatty, C. Nsoedo, and D. Parker, "Human Factors in Anaesthetic Practice: Insights from a Task Analysis," *Br. J. Anaesth.*, vol. 100, no. 3, pp. 333–344, 2008.

[29] G. van der Veer, B. Lenting, and B. Bergevoet, "GTA: Groupware Task Analysis - Modeling Complexity," *Acta Psychol. (Amst).*, vol. 91, pp. 297–322, 1996.

[30] B. Kirwan and L. Ainsworth, Eds., *A Guide to Task Analysis*. CRC Press, 1992.

[31] D. Embrey, "Task Analysis Techniques," 2000. [Online]. Available: http://www.humanreliability.com/articles/Task Analysis Techniques.pdf.

[32] R. Clark, D. Feldon, J. Merrienboer, K. Yates, and S. Early, "Cognitive Task Analysis," in *Handbook of research on educational communications and technology*, 2008, pp. 577–593.

[33] S. Card, T. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, 1983.

[34] B. John and D. Kieras, "The GOMS Family of User Interface Analysis Techniques: Comparison and Contrast," *ACM Trans. Comput. Interact.*, vol. 3, no. 4, pp. 320–351, 1996.

[35] M. Matsubara, H. Terasawa, H. Kadone, K. Suzuki, and S. Makino, "Sonification of Muscular Activity in Human Movements Using the Temporal Patterns in EMG," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2012.

FROM PROGRAM MUSIC TO SONIFICATION: REPRESENTATION AND THE EVOLUTION OF MUSIC AND LANGUAGE

Stephen Taylor

School of Music, University of Illinois
1114 W. Nevada St. Urbana, IL 61801, USA
staylor7@illinois.edu

ABSTRACT

Research into the origins of music and language can shed new light on musical representation, including program music and more recent incarnations such as data sonification. Although sonification and program music have different aims—one scientific explication, the other artistic expression—similar techniques, relying on human and animal biology, cognition, and culture, underlie both. Examples include Western composers such as Beethoven and Berlioz, to more recent figures like Messiaen, Stockhausen and Tom Johnson, as well as music theory, semiotics, biology, and data sonifications by myself and others. The common thread connecting these diverse examples is the use of human musicality, in the bio-musicological sense, for representation. Links between musicality and representation—dimensions like high/low, long/short, near/far, etc., bridging the real and abstract—can prove useful for researchers, sound designers, and composers.

1. INTRODUCTION

The emerging field of bio-musicology [1] and research into the origins of music and language [2], [3] can shed new light on musical representation, including program music and more recent incarnations such as data sonification. Although sonification and program music have different aims—one scientific explication, the other artistic expression—similar techniques, relying on human and animal biology, cognition, and culture, underlie both.

Two of the earliest and most successful examples of sonification, the Geiger counter (1928) and Morse code (1836), are still widely used today. The Geiger counter, a radiation detector, emits a series of clicks: the faster the clicks, the greater the danger [4]. It capitalizes on instincts shared by humans and primates: the chimpanzee pant-hoot follows the same rising curve of acceleration and intensity, as do musical topics such as the Mannheim rocket (the term comes from 18th-century symphonies that begin with an ascending arpeggio and crescendo).

Morse code on the other hand is abstract but general. While it is unintelligible unless you know the code, it can encode literally any message; it is a kind of musical cryptography, similar to Bach's use of gematria. These sonic techniques recall two recent hypotheses for the evolution of music. One [5] holds that musicality is innate, encoded in our genes. The other sees music as a "transformative technology of the mind," akin to the control of fire [3]. Although the knowledge of fire is not built into our DNA, it has nonetheless profoundly affected our culture, our bodies, our biology.

Human musicality likely comprises both of these ideas, and others as well, as Bruno Nettl writes: "…I have become convinced that the things we call music began in a number—maybe a lot—of different ways, some going back beyond the evolution of homo sapiens. Some are no doubt older than others, but the 'younger' ones did not necessarily develop from older ones." [6]. W. Tecumseh Fitch, similarly, advocates for the study of bio-musicology: "the biological study of musicality in all its forms… While music, the product of human musicality, is extremely diverse,
musicality itself is a stable aspect of our biology and thus can be productively studied from comparative, neural, developmental and cognitive perspectives" [1]. This essay explores the continuum from innate musicality to learned, symbolic representation; or, from music that mimics extra-musical things, to music built on extra-musical information.

2. ICON, INDEX, SYMBOL

Morse code and the Geiger counter can also be understood using the semiotic triad of icon-index-symbol, developed by the American Charles Peirce (1839-1914; see also Turino [7] for a detailed discussion from a musical perspective). The Geiger counter is indexical (the greater the radiation, the greater the speed of the clicks); Morse code, by its assignment of letters to patterns of dots and dashes, is symbolic. (For the English codebook, each pattern is chosen for how often its letter appears; while the result might sound arbitrary, careful listening reveals which signals occur most often.) Another early sonification device, the stethoscope, can be interpreted as iconic. Dombois and Eckel consider the stethoscope a kind of audification: "…one of the few important examples of an accepted scientific device using audio" [8].

The quasi-arbitrary, symbolic nature of Morse code is shared with spoken language, as Fitch notes in his 2010 book *The Evolution of Language:* "…arbitrariness is almost automatic if you start with a vocal system, for the realm of the iconic is rather limited in vocalizations. Onomatopoeia can buy you some animal names, and some emotional expressions, via imitation, but not much more. But the flip side of the coin—too often overlooked—is that arbitrariness is a crucial step to a fully open field for semantic reference, and this is something that we gain almost automatically with the capacity to link meanings to vocal signals..." [9], p. 467.

Seen from this semiotic perspective of icon, index, and symbol, program music and sonification both span a

https://doi.org/10.21785/icad2017.060

continuum (Figure 1), which is indebted to Kramer's classification of sonification from analogic to symbolic [10].

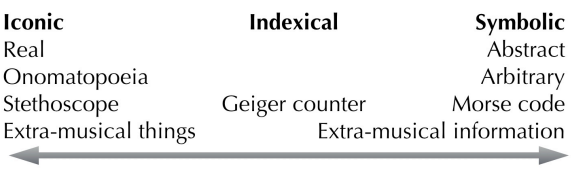


Figure 1. Sonification and program music both span a continuum from the real to the abstract.

Just as language and sonification range from onomatopoeia to arbitrary symbols, so does program music. Many composers have exploited the timpani's resemblance to thunder, most notably Beethoven in his storm movement from the Symphony no. 6. At the opposite end of the continuum, the *idée fixe* in Berlioz's *Symphonie Fantastique* is an abstract symbol, a musical theme representing the object of his infatuation (Wagner's use of leitmotif works in much the same way: not only are there themes for different characters, there are also themes that represent abstract concepts, say, the renunciation of love). In this sense, sonification can be thought of as an updated version, or generalization, of program music.

3. PROGRAM MUSIC

Composers have explored this continuum from the real to the abstract for centuries. An early instance of program music is Marin Marais' "Le tableau de l'Operation de la taille", about kidney stone surgery (Figure 2).

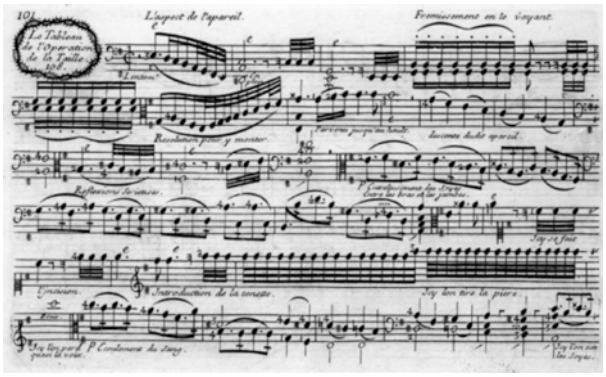


Figure 2. Marin Marais (1656-1728), "Le Tableau de l'Operation de la taille" for viol; an early instance of program music [11].

This work resembles many of Marais' other works for viol. Without the accompanying explanations ("Appearance of the device"; "Here is the incision"; etc.), a listener might not have any idea that the piece is about kidney stone surgery (although Marais does indulge in word painting: e.g. "Descent of the device" at the end of the second line is depicted by a slow, descending scale). But the form of the piece is unusually choppy, jumping suddenly from one musical idea to another; the short explanations scattered throughout dictate the sudden changes, thwarting the music's formal cohesion. In fact the piece is almost always performed with the words spoken as narration: the music becomes a kind of illustration of the text, a historical precedent to works like Prokofiev's *Peter and the Wolf*. The explanations are a kind of caption, and they raise questions: is music somehow less "valid" if it can only be understood via a caption or program? Does a caption's presence somehow obviate the music's role, to "sound like" the thing or information it's representing? Can music be simultaneously abstract and descriptive? How to represent something, caption or not, that lacks a sonic analogy in the real world? We will come back to these questions, and the issue of captions.

I have already mentioned the thunderstorm in Beethoven's Symphony no. 6, but the second movement (Figure 3), "Scene by a brook", is just as evocative. At the end of the movement the orchestra drops out, leaving a single flute, who starts a trill; Beethoven's sketch shows that this is a nightingale. An oboe joins in (quail), followed by a clarinet (cuckoo). At this moment—among the most famous passages in the history of program music—a particularly striking feature is the interval of the descending third. Beethoven chose it carefully. When highlighted by the solo clarinet, it becomes a cuckoo (and at least for me, this is the most "bird-like" of the three); but for the entire movement, the interval has been embedded in the flowing accompaniment. It is as if Beethoven uses the same musical fabric to weave both an abstract design and a vivid portrait.

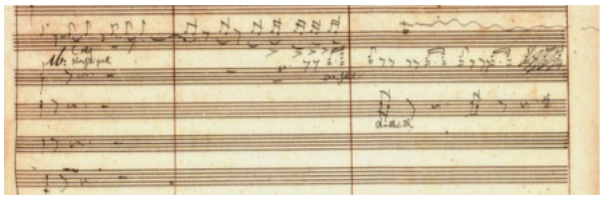


Figure 3. Beethoven, passage with birds from Symphony no. 6, II. "Szene am Bach" [12]. An example of aural mimicry, or icon; the birds are nightingale, quail, cuckoo. The cuckoo's interval of the third is also embedded within the movement's flowing accompaniment.

We use language in a similar way, as Fitch describes above, when we use onomatopoeia in a sentence: spoken language can use the same sound in an iconic or symbolic way, depending on context. In his 2005 book *The Singing Neanderthals,* Steven Mithen discusses several studies in which onomatopoeia plays a role in non-obvious subjects ([2], p. 170). As he describes, in the 1920s Edward Sapir "undertook an intriguing and quite simple test. He made up two nonsense words, *mil* and *mal,* and told his subjects that these were the names of tables. He then asked them which name indicated the larger table and found… that almost all of them chose *mal.*" (As an exception that proves the rule, the writer David Foster Wallace kept lists of words which, counter-intuitively, sounded like the opposite of their meaning; one of his favorites was "pulchritudinous" [13].)

Mithen speculates that these kinds of sounds may have played a role in the evolution of language. Beethoven, again in sketches for the Sixth Symphony, argues for a similar idea in music (Figure 4). Both in language and music, this kind of musical mimicry can be considered as indexical in the Peircean sense; as the sound changes, so does the thing it describes.

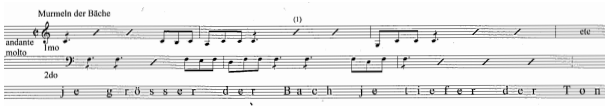


Figure 4. From Beethoven's Eroica sketchbook: "The bigger the stream, the deeper the tone" [12]. This phrase can be interpreted as indexical, in the Peircean sense (in nature, bigger things are associated with lower frequencies).

There are far too many other examples of musical onomatopoeia to list, but we must mention Olivier Messiaen's magnificent depiction of birds in works like *Oiseaux éxotiques* for piano and wind orchestra (1959), and the massive *Catalogue d'oiseaux* for solo piano (1956-58). These can be heard as a kind of updated version of Beethoven's birds, famously realistic, although sometimes they are slowed down and distorted to the point of unintelligibility, invoking gigantic, imaginary creatures [14]. Today with modern sampling technology it is easy to use recorded sounds from anywhere, both as an iconic reference and as an abstract element embedded into the music (like Beethoven's thirds). Two examples illustrate this point: Debussy evokes the feeling of walking in the snow with his piano prelude "Des pas sur la neige" (1909-10), while Björk, on her song "Aurora" from the 2001 album *Vespertine,* uses a Foley-like sample of someone actually walking in the snow to create the song's percussion (performed in concert by a live snow-walker).

Finally, despite my emphasis on musical mimicry, it is important to recognize that this notion has not gone unchallenged. Werner Wolf [15] notes that in English we have words for description (writing), and depiction (visualizing), but we have no verb "to desound". He continues: "[Music] is the most **abstract and non-referential medium** of all the arts and media, and it is therefore sometimes claimed that a piece of music does not consist of signs at all, in other words that music has no semiotic quality like verbal language… One should, however, be more precise, for music can be said to be 'referential', but mainly in the sense of '**self**-referential' rather than of '**hetero**-referential'. The reason for this is that music consists mainly of signs whose signification resides in their ability to point to other signifiers within the same system, usually by iconically imitating or repeating them (but also by forming contrasts to them)" (his emphasis, p. 59; canons are good examples of this self-referential quality).

Franz Liszt, quoted by Roger Scruton in the *New Grove Dictionary,* moderates this view: he did not "regard music as a direct means of describing objects; rather he thought that music could put the listener in the same frame of mind as could the objects themselves" [16]. Berlioz, in his groundbreaking essay "On imitation in music" [17], confronts this kind of thinking head on: "The famous naturalist Lacépède … says somewhere that 'since music has only sounds at its disposal, it can act only through sound. Hence in order to produce the signs of our perceptions these signs must themselves be sounds.' But how can one express musically things that make no sound whatever, such as the denseness of a forest, the coolness of a meadow, the progress of the moon? Lacépède answers, 'By retracing the feelings these things inspire in us.'… I am far from sharing that opinion … Is there, for example, any single fixed manner in which we are affected by the sight of a forest, a meadow, or the moon in the sky? Assuredly not" (p. 43-44). Rather than "retracing the feelings these things inspire in us", which Berlioz notes is hopelessly subjective, I will argue that we can find ways of representing information that, even if not iconic, follow paths laid down by our innate sense of bio-musicality, shared among humans and other animals.

4. TOPIC, GESTURE, AND INDEX

Keeping these caveats in mind, but moving nevertheless along the continuum from onomatopoeia, we find musical topoi and gestures [18], [19]. A common example of a topic is the march, which bears an obvious relation to walking: the duple meter reflects our bipedal nature (guitarist Mark Stewart in a 2017 personal communication wonders if alien life, or even the octopus, could have different musical meters). Topics are not necessarily iconic, in the Peircean sense. Maybe the march can be considered a kind of index: the faster the march, the faster one marches. Other topoi include dance music, fanfare, lament, serenade, lullaby, etc. From the bio-musicological perspective, musical topoi go quite deep: Brown and Jordania, in their list of musical universals, note that "Music-induced emotions vary widely, from arousing (e.g., marching music) to soothing (e.g., lullabies…)" ([5], p. 240).

Gesture is a more general concept than topic, and harder to pin down. Hatten [19] defines a musical gesture as a "perceptible and significant energetic [intensity] shaping [frequency, timbre] through time [duration], regardless of modality or channel" (p. 108). These changing energies again recall the idea of an index, which also changes through time. Gestures can allow for greater subtlety than topics (although composers skillfully combine topics to create emotional nuance): "How, in other words, might one go beyond the major versus minor, happy versus sad correlation, when there are more complex expressive meanings at work?" (p. 13).

Ascending and descending gestures are common in music, language and beyond (see the Mannheim rocket and chimpanzee pant-hoot mentioned above), but they raise an interesting "polarity" problem, as noted by Grond and Berger, [20]: "When one of the authors' daughter started studying the 'cello she confused pitch direction and the verbal descriptions of 'higher' and 'lower'" (p. 385); the cellist must move their arm lower down the fingerboard to produce a higher pitch. Barrass and Vickers [4], in the same volume, also describe sonification experiments on subjects with impaired vision, who don't necessarily use the words "higher" and "lower" to describe pitch in the usual manner (p. 148).

Although we may not all use the words "higher" and "lower" in the same way, we (humans and other animals) have similar reactions to higher and lower frequencies. Animal researchers have studied emotional communication among primate young who are temporarily separated from their mothers [21]. The authors acknowledge that experiments like this cannot be done with human subjects—but they find a surprising bio-musicological relationship with opera. "Duets in which the partner addressed is in sight or approaching and subsequently a unification of separated partners are not yet studied in human real-life scenarios but can be found in numerous reunion scenes in operas. These duets start by increased frequency of alternating interjections ('vocal rate'), increased pitch, loudness, and highly modulated rising pitch

contour. Subsequently, a duet, symbolizing the unification of separated partners follows. Obvious examples for such a sequence of vocalizations can be found in the operas *The Magic Flute* by Mozart, *Fidelio* by Beethoven, *Othello* by Verdi, *Carmen* by Bizet, and *Three Penny Opera* by Brecht/Weill…" (p. 347). These gestural changes in pitch, loudness etc. are indexical, analogous to changing emotions and arousal: not only for humans but for other mammals as well. A response which runs that deeply within us should be able to be put to use for other purposes, to represent a variety of phenomena. In this sense, sonification can use our bio-musical instincts in a way that recalls Stephen Jay Gould's "spandrels": an evolutionary adaptation, co-opted for another purpose.

Composers in the 20th century experimented with other ways of using extra-musical phenomena as musical indices. Heitor Villa-Lobos (*New York sky line* for piano, 1957) and others converted the New York City skyline into a melody, by mapping it onto staff paper (Figure 5). Here we begin to approach the idea of sonification, the representation of data as non-speech sound; or, if you prefer, data-driven music. We can't say that Villa-Lobos is "mimicking" the skyline, because the skyline doesn't make any sound: instead he is using the skyline as an index to musical pitch. (Earlier composers such as Bach would notate melodies and fugue subjects to represent the Cross; this kind of orthography goes back at least into the Renaissance.) As noted in the caption to Figure 5, this kind of reference can be seen as a gimmick, and indeed has been for centuries. In a letter to his parents, Mendelssohn [22] complained bitterly about the sensational, programmatic quality of Berlioz's *Symphonie Fantastique:* "How utterly loathsome this is to me, I don't have to tell you. To see one's most cherished ideas debased and expressed in perverted caricatures would enrage anyone. And yet this is only the program. The execution is still more miserable: nowhere a spark, no warmth, utter foolishness, contrived passion represented through every possible orchestral means…" But it must be said that the Villa-Lobos is quite beautiful: the skill of the translator matters greatly when converting data to music (also see Kramer [10]: "The craft of composition is important to auditory display design").

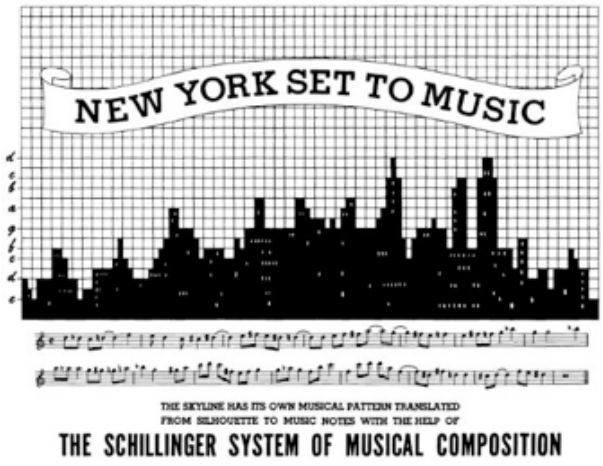


Figure 5. Heitor Villa-Lobos and others have used the New York City skyline for musical compositions, an example of data-driven music. Note the possibility (or danger) for sensationalism and gimmickry.

Karlheinz Stockhausen uses a more subtle approach to indexicality in his work *Gruppen* for three orchestras (1955-57). As he describes in his article "How Time Passes" [23], he looked at the mountains from his window in Switzerland, and traced their contour to provide the timbres ("formant-spectra", Figure 6) for his instrumental forces. No less sensational perhaps, but not as directly audible to the listener—which raises problems for the researcher who wants to communicate data as clearly as possible. (Regarding the Stockhausen, I would argue that mapping the vertical y-axis to timbre does not make the best bio-musicological use of innate musicality; a dimension that captures the relation of dark to light, or near to far, may be more suitable for timbre.)

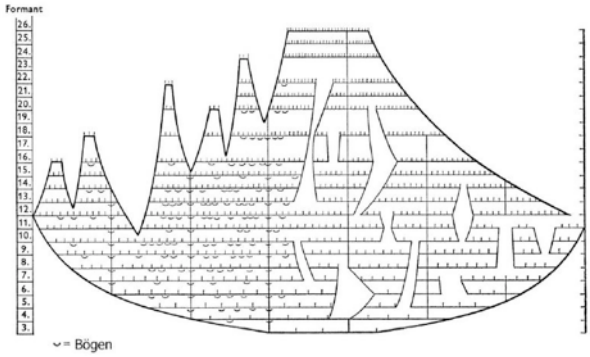


Figure 6. Karlheinz Stockhausen traced the contours of the Alps to provide the instrumentation for part of his composition *Gruppen;* an example of timbre controlled by extra-musical data.

As we saw with the Marais in Figure 2, this problem of communication can be addressed by a caption. In essence, the "program" in program music is a kind of caption that informs the listener about what they're hearing, just like a caption for a graph or chart. Without its caption, a graph is an abstract design (think of the London Underground map without labels). Berlioz [17] also weighs in on the idea of caption: "…it is strictly required that the hearer be notified of the composer's intent by some indirect means, and that the point of the comparison be patent. Thus Rossini is thought to have depicted in *William Tell* the movement of men rowing. In point of fact all he has done is to mark in the orchestra a *rinforzando* accented at regular intervals—an image of the rhythmic straining of the oarsmen, whose arrival has been announced by the other characters."

The Paris-based composer Tom Johnson has come up with an ingenious way to incorporate captions in works like *Bedtime Stories No. 12* (1986), based on the stock market; and *Narayana's Cows* (1989), based on an infinite series discovered by the 14th-century Indian mathematician (Figure 7). In these pieces, a narrator provides a spoken caption between each bar of music, explaining what the audience is about to hear. Johnson describes his approach to captions in the preface to *Narayana's Cows*: "The text is neither musical analysis, nor a math lesson, nor comic relief. It should be delivered simply and directly as an integral part of the piece, either by the musicians or by someone else." Higher, longer notes represent a mother cow; lower, shorter notes represent the mother's calves, in an indexical relationship. The harmony mirrors the rhythmic mapping: the hexatonic mode on which the piece is based alternates between two unequal intervals, minor third and minor second. Thanks to the

narration, the relationship is so obvious that it's almost iconic. This obviousness in the wrong hands could lead to a cartoon-like caricature, as Mendelssohn complains about Berlioz. As with the Villa-Lobos, it is a measure of Johnson's musicianship, inventiveness and taste that the work is so successful.

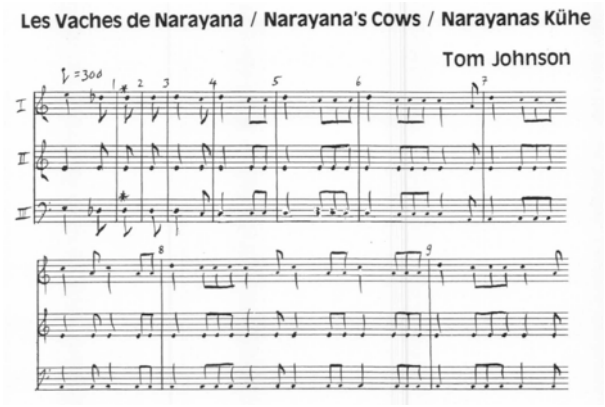


Figure 7. Tom Johnson's *Narayana's Cows* (1989) uses a narrator to provide a spoken caption between each bar; after each caption the ensemble always restarts from the beginning and plays up to the next bar.

This kind of spoken caption for musical works is rare: besides the Marais, earlier examples include Prokofiev's *Peter and the Wolf* (1936) and Britten's *Young Person's Guide to the Orchestra* (1945), both composed for children (which makes Johnson stand out even more, as a contemporary composer who works with captions for general audiences). Another example of caption in Johnson's earlier work is *Failing: a very difficult piece for solo string bass* (1976), in which the performer must recite a running philosophical commentary while attempting to play the piece (which eventually becomes impossible, hence the title). Other recent examples include sonifications or sonic illustrations in radio shows and podcasts, such as Radiolab's 2015 episode "Antibodies Part 1: CRISPR [24].

5. SYMBOLIC SOUND

As we approach the abstract, symbolic end of the continuum shown in Figure 1, we encounter musical analogues to Morse code: generalized, more or less arbitrary sonic symbols that can convey any desired meaning (of course, this description also applies to spoken language). Perhaps the most well known of these is musical cryptography, in which musical notes stand for letters of the alphabet or other symbols. Examples include BACH (B-flat, A, C, B-natural), and Dmitri Shostakovich's signature motive DSCH (D, E-flat, C, B-natural). Gematria, the mystical practice of assigning numbers to letters of the alphabet (e.g. 666, the number of the beast in the Book of Revelation), has also long been practiced by composers, including Bach.

Messiaen's "communicable language" provides a more recent example. Instead of single pitches, he uses pungent chords to form a sonic alphabet, spelling out messages from the Hebrew Bible [25]. One criticism of Messiaen's system is that, like language, its meaning depends on the arbitrary assignment of meaning to different sounds. Unless you know Messiaen's system (or you have access to the score, where he labels each chord with its corresponding alphabet letter), there is no way to discern the sacred texts. Julian Anderson pushes back against this criticism: "Any alert listener, even if unaware of the composer's detailed intentions, will not fail to register in Messiaen's most characteristic pieces the repeated impression of vivid musical signals being given forth as declaimed utterances one after the other, usually grouped into the distinct juxtaposed blocks so typical of his mature work" (p. 257). In this way, perhaps, Messiaen is using his invented alphabet as a spur to his creative impulse, a new means of creating music in his own style; not so different than Berg's use of serialism, manipulating tone rows for his own expressive, tonal, emotional goals; or perhaps Stockhausen, tracing the Alps to get new ideas for orchestral timbre.

This idea of an arbitrary conveyor of meaning, dependent on a kind of translation from music to language, recalls the sonification of data. Barrass [26] describes an aesthetic continuum, or teeter-totter, similar to Figure 1, but with music at one end and sonification at the other: "The intention to produce a musical experience does not necessarily include the intention to reveal explicit information about the sources of composition. However, when the composer does intend the listener to understand extra-musical information, the work then enters the realm of sonification." (p. 146).

This is an intriguing idea, to pit the musical material (derived from data or an algorithm) against the musical experience, to decide whether something is sonification or music. It makes sense, if we think back to the "unmusical", choppy form of the Marais kidney-stone piece; here the data is the narrative of the surgery. This kind of formal choppiness is part of what people find cheap about cartoonish musical caricatures: it is as if the music is subservient to an outside driver, rather than following its own abstract, non-representational course. (An important exception to this dichotomy, though, would be the minimal music of Steve Reich (e.g. *Piano Phase,* 1967), in which the self-referential, canonic algorithm revealing itself *is* the musical experience.) *Narayana's Cows* seems to perch exactly on the tipping point between music and sonification; it could go either way, depending on how you listen.

6. BIO-MUSICOLOGY AND SONIFICATION

Returning to bio-musicology, if musicality is innate, then there could exist intuitive (or at least, not completely arbitrary) ways of representing even abstract data. Some of these possibilities are suggested by Kofi Agawu [18] in a series of oppositions: "…the so-called binary classification, in which the relationships between phenomena are perceived as oppositions, may also be seen in the metaphors that we apply to various dimensional behaviors: pitch and register are conceptualized within a high-low axis, rhythm and duration on a long-short axis, timbre on a dark-bright axis, texture on a thick-thin axis, and so on." By applying these metaphors (and others) to extra-musical information, we can find ways of representing complex data that listeners can understand more intuitively. Fitch [1] describes these metaphors as "a comparative approach, which seeks and investigates animal homologues or analogues of specific components of musicality, wherever they can be found." (Nettl [6] cautions us to "be careful in transferring the labels of human taxonomies—of Western taxonomies, actually—too readily to other species.")

This kind of metaphorical thinking is closely related to Grond and Berger's work on Parameter Mapping Sonification [20]. As they admit, "the lack of standards and ubiquity in mapping strategies often makes sonification research akin to working on the tower of Babel" (p. 387). But at the same time, "Effective sonification must be intuitive and easily learned" (p. 388). By applying some long-held ideas from composers working with program music, as well as principles of bio-musicology, researchers can meet these challenges, and perhaps arrive at some standards for mapping sound to data. These are simple and intuitive: high/low, long/short, near/far, etc. Much of the subtlety of rhythm and harmony—what Barrass calls the "musical experience"—might get lost. The goal is to keep the basic dimensions simple and recognizable, while using musical aspects like rhythmic and harmonic subtlety to heighten the aesthetic quality of the sonification (although there may be ways to use these subtleties in a parameter mapping). Graphical analogies can be found in works like Edward Tufte's *The visual display of quantitative information* [27].

My final example, an attempt at using multiple dimensions of sound simultaneously to represent corresponding data dimensions, is a 30-second video of protein folding (screenshot, Figure 8). The goal is to present dimensions of data, sonically, that are orthogonal to the animated computer model on the right.

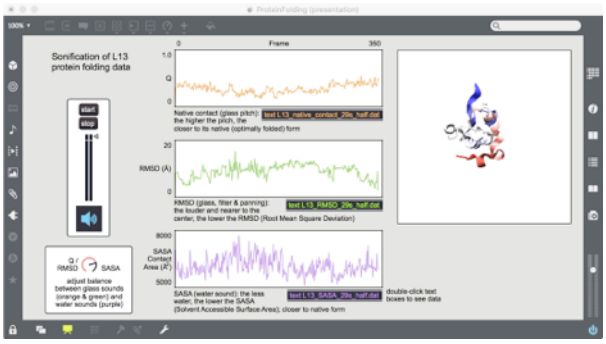


Figure 8. Sonification of protein folding by the author, made with Max; video and download links can be found at www.stephenandrewtaylor.net/genetics.html. A 30-second animation of the rapidly shifting protein plays on the right (video animation courtesy of Martin Gruebele). In the middle are shown three orthogonal data dimensions, explained in the text.

In the middle are shown three graphs representing different dimensions of data which cannot be portrayed by the video animation. The orange line at top represents native contact, or how close the protein is to its optimal shape; the green line below it shows RMSD (root mean square deviation), or how far away the protein is from its native form; the purple line at bottom shows SASA (solvent accessible surface area, or "leakiness"). Each is sonified in a different way, using a different dimension (or axis in Agawu's words). The orange line is played by a percussive glass sound, where pitch height is on the y-axis; the higher the pitch, the closer to the protein's ideal form (since the folding data are sampled at a constant speed, rhythm is constant). The green line (deviation from ideal form) *modulates* the orange line: the higher the green line, the more the glass sound is filtered and panned; this makes it sound farther away, both in distance and in the stereo field. When the glass sound (the orange line) sounds very close and centered, then the deviation is low. By combining these data dimensions with the video animation on the right, it is possible to hear the data with caption-like visual reinforcement (the three graphs in the middle). By focusing visually on the protein animation while listening, one can simultaneously perceive multiple data streams.

One note on the aesthetics of the glass sound: in my previous sonification attempts, I have been frustrated by the artificial quality of MIDI and synthesis. So for this example, I recorded several different percussive wine glass sounds (gently striking the glass with a chopstick). For each glass sound (all coming very fast, 12 notes per second, with 24 video frames per second), a Max patch randomly selects one of the wineglass sounds. The result sounds more like someone actually playing an instrument, which contributes to the aesthetic quality (I tried various synthesis options as well, but at least to my ears none of them sounded as good as using recordings).

Finally, the purple line, showing the "leakiness" of the protein, is represented by water droplet sounds (also chosen randomly from about a dozen samples, all coming very fast). The higher the line, the louder the droplet. A rain sound is also constantly present, representing a smoothed version of the purple graph. Because the "amount" of rain depends on the y-axis, I cannot use an existing recording of rain; so it is synthesized, following techniques outlined by Andy Farnell [28].

By using musical sound (or iconic *musique concrète* in the case of the water drops and rain sounds) to represent three data dimensions in addition to the video animation, the presentation is more informative, and arguably more effective. The sonification technique uses both icon (water) and index (the up-down axis and the near-far axis). The glass percussion sound is a symbolic representation of the protein's shape as it rapidly changes. Each of these dimensions, of course, requires a caption to make sense to the listener, no different than Marais or Berlioz. It can interpreted either as music or sonification, as Barrass points out [26], depending on whether the listener focuses on the "musical experience" or the "extra-musical information", although I have designed it more on the sonification end of the continuum (if anything, it sounds a little like Rimsky-Korsakov's "Flight of the bumblebee", itself an evocative, buzzing portrait, hovering between index and icon).

7. CONCLUSION

As a composer I have long been inspired by science, and written many works of "program music" inspired by scientific phenomena. Over the past several years I have grown dissatisfied with this approach. Just as Messiaen's birds are much more faithful to reality [14] than Beethoven's, composers are finding it is possible to create a new kind of program music that is actually built on extramusical information, not just inspired by it. Data itself is inspiring. And as we learn more about the origins of music and language, and the nature of human and animal musicality, we can learn to portray this data more effectively, more intuitively.

8. ACKNOWLEDGEMENTS

I am grateful to the Guggenheim Foundation for a 2014-15 grant providing time to dive into sonification, an area I had been interested in for a long time but never studied closely enough. I'm also grateful to Bruce Walker and John Hummel for conversations in 2015 that pointed me in the right direction. Thanks to Bruno Nettl and the anonymous reviewers of an earlier draft of this essay; to William Kinderman for showing me the Beethoven examples; to Dmitri Tymoczko for introducing me to *Narayana's Cows;* to Daniel Stelzer for help with coding; to Martin Gruebele and Zhang Yi for the protein folding video and data; and to Carla Scaletti for inventing the phrase "symbolic sound".

9. REFERENCES

[1] W. T. Fitch, "Four principles of bio-musicology," in *Philosophical Transactions B,* vol. 370, no. 1664, March 2015.

[2] S. Mithen, *The singing Neanderthals: the origins of music, language, mind and body,* London, UK: Weidenfeld & Nicolson, 2005.

[3] A. Patel, *Music, language, and the brain,* New York, USA: Oxford University Press, 2008.

[4] S. Barrass and P. Vickers, "Sonification design and aesthetics," in *The sonification handbook,* ed. Hermann, Hunt, Neuhoff. Berlin, Germany: Logos Verlag, 2011, pp. 145-172.

[5] S. Brown and J. Jordania, "Universals in the world's musics," in *Psychology of Music,* vol. 41, no. 1, pp. 229-248, 2013.

[6] B. Nettl, "Have you changed your mind? Reflections on sixty years," Champaign IL, USA: Elephant & Cat, 2016.

[7] T. Turino, *Music as social life: the politics of participation,* Chicago IL, USA: University of Chicago Press, 2008.

[8] F. Dombois and G. Eckel, "Audification," in *The sonification handbook,* ed. Hermann, Hunt, Neuhoff. Berlin, Germany: Logos Verlag, 2011, pp. 301-324.

[9] W. T. Fitch, *The evolution of language,* Cambridge, UK: Cambridge University Press, 2010.

[10] G. Kramer, "An introduction to auditory display," in *Auditory display: Sonification, audification, and auditory interfaces,* ed. Kramer. Reading, MA, USA: Addison Wesley, 1994, pp. 1-78.

[11] http://imslp.org

[12] L. v. Beethoven, *Beethoven's "Eroica" sketchbook: a critical edition,* ed. Lockwood and Gosman. Champaign IL, USA: University of Illinois Press, 2013.

[13] D. T. Max, *Every love story is a ghost story: a life of David Foster Wallace,* New York, USA: Granta Publications Ltd., 2012.

[14] R. Fallon, "The record of realism in Messiaen's bird style," in *Olivier Messiaen: Music, art and literature,* ed. Dingle and Simeone. London, UK: Ashgate Press, 2007, pp. 115-136.

[15] W. Wolf, "Description as a transmedial mode of representation: general features and possibilities of realization in painting, fiction and music," in *Description in literature and other media,* ed. Wolf and Bernhart. Amsterdam, The Netherlands: Rodopi B.V., 2007, pp. 1-90.

[16] R. Scruton, "Programme music," in *The New Grove dictionary of music and musicians,* London, UK: Oxford University Press, 2001, p. 396.

[17] H. Berlioz, "On imitation in music," trans. J. Barzun in *Fantastic Symphony: A critical edition,* ed. Cone. New York, USA: W.W. Norton & Co., 1971). Orig. pub. In *Revue et Gazette musicale,* January 1 and 8, 1837.

[18] K. Agawu, *Playing with signs: a semiotic interpretation of classic music,* Princeton NJ, USA: Princeton University Press, 1991.

[19] R. Hatten, *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, Schubert,* Bloomington IN, USA: Indiana University Press, 2004.

[20] F. Grond, J. Berger, "Parameter mapping sonification," in *The sonification handbook,* ed. Hermann, Hunt, Neuhoff. Berlin, Germany: Logos Verlag, 2011, pp. 363-398.

[21] E. Altenmüller, S. Schmidt, and E. Zimmerann, "A cross-taxa concept of emotion in acoustic communication: An ethological perspective," in *Evolution of emotional communication: From sounds in nonhuman mammals to speech and music in man,* ed. Altenmüller, Schmidt, Zimmerman. London, UK: Oxford University Press, 2013, pp. 339-355.

[22] F. Mendelssohn, from a letter to his mother, Rome, March 15, 1831, in *Composers on music: Eight centuries of writings,* ed. Fisk. Boston MA, USA: Northeastern University Press, 1956. p. 85.

[23] K. Stockhausen, "How time passes," in *Die Reihe,* Vienna, Austria: Universal Edition, vol. 3, pp. 13-42, 1957.

[24] http://www.radiolab.org/story/antibodies-part-1-crispr/

[25] J. Anderson, "Messiaen and the problem of communication," in *Messiaen perspectives 1,* ed. Dingle and Fallon. Surrey, UK: Ashgate Publishing Limited, 2013, pp. 257-268.

[26] S. Barrass, "The aesthetic turn in sonification towards a social and cultural medium," in *AI & Society*, London, UK: Springer-Verlag, vol. 27, pp. 177-181, 2012.

[27] E. Tufte, *The visual display of quantitative information,* Cheshire CT, USA: Graphics Press, 1992.

[28] A. Farnell, *Designing sound,* Cambridge MA, USA: MIT Press, 2010.

Paper Session 2
Movement

MUSICAL EXPECTANCY IN SQUAT SONIFICATION FOR PEOPLE WHO STRUGGLE WITH PHYSICAL ACTIVITY

Joseph W Newbold & Nadia Bianchi-Berthouze

UCLIC
UCL
London, UK
joseph.newbold.14/n.berthouze@ucl.ac.uk

Nicolas E. Gold

Dept. of Computer Science
UCL
London, UK
n.gold@ucl.ac.uk

ABSTRACT

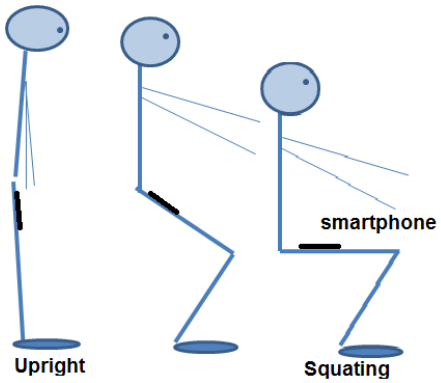
Physical activity is important for a healthy lifestyle. However, it can be hard to stay engaged with exercise and this can often lead to avoidance. Sonification has been used to support physical activity through the optimisation/correction of movement. Though previous work has shown how sonification can improve movement execution and motivation, the specific mechanisms of motivation have yet to be investigated in the context of challenging exercises. We investigate the role of music expectancy as a way to leverage people's implicit and embodied understanding of music within movement sonification to provide information on technique while also motivating continuation of movement and rewarding its completion. The paper presents two studies showing how this musically-informed sonification can be used to support the squat movement. The results show how musical expectancy impacted people's perception of their own movement, in terms of reward, motivation and movement behaviour and the way in which they moved.

1. INTRODUCTION

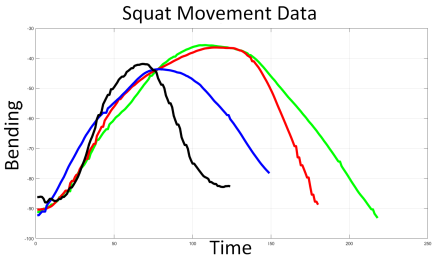
Despite the majority of adults having a positive attitude toward physical activity, most do not meet the minimum UK government recommendation for physical activity [1] and adherence to physical activity is poor [2]. Prior work has shown how in-the-moment feedback can be used to optimise and improve the quality of movement [3, 4, 5, 6, 7]. People who struggle with physical activity often not only struggle with the form and technique of a specific movement but also with psychological barriers related to the perceived capability to perform such activity [8]. In this paper, we investigate how real-time movement sonification enriched with musical expectancy affects people's motivation during a challenging physical exercise.

The current use of sonification in physical activity primarily only informs users of their movement and does not focus on the specific mechanism within sound that can support people overcoming psychological barriers to physical activity [3, 4, 5]. Conversely, literature on behavioural changes has shown the importance of addressing psychological barriers to motivate an increased general activity level [9, 10, 11, 12]. However, such studies have focused only on long-term changes in the amount of activity by typically using goal-setting, quantify-self and post-activity rewards and reflection type of approaches. They fail to address the psychological barriers to performing a challenging movement.

In this paper, we aim to combine these two ideas: a real-time sonification during the performance of a challenging exercise which provides in-the-moment motivation and reward. To achieve this, we combine the use of auditory display (which has been shown to aid motor learning and improvement [3, 4, 5]) and principles from music theory and cognition which have an impact on not only people's emotional state [13] but also on the embodied perception of one's movement [14, 15]. We focus on the principle of musical expectancy (the way we expect a piece of music to con-



(a) Set up for measuring the squat movement



(b) The movement data for one participant's first squat in Study one

Figure 1: a) The phone on the the upper leg measures the angle between the leg and the floor. b) The movement data measured by the phone: Green-musically stable, Red-musically unstable, blue-non-stability and black-no sound

https://doi.org/10.21785/icad2017.008

tinue) to affect the way people endure in reaching the ending point of a movement.

We build on the work by Newbold et.al that showed how different harmonic stability at the target point of a movement can be used to motivate progress and reward the completion of movement [16]. However, that work focused on the stretch forward movement, a movement which has no defined ending to the movement and with a population that did not struggle with the performance of that movement. It has yet to be seen how this kind of sonification may differ from a more challenging exercise and one that has stronger perceptual (proprioceptive) cues at the target point. We focus on the squat down movement because 1) it is a fundamental movement in exercise [17] but where people often struggle with; and 2) it is one in which it is difficult to assess one's own form visually and is often performed incorrectly [17, 18]. Specifically, we focus on the depth of the squat movement, which is often misjudged by people who struggle with squats [17]. This study, therefore, aims to examine how this kind of movement may be impacted by different types of expectation at the ending of the movement sonification. We report here two studies that extend Newbold et al.'s work to investigate further the mechanisms surround musical expectancy within sonification. The first study aims to understand the effect of expectancy over motivation and movement within the squat movement by comparing it with a sonification that does not carry any expectancy information at the termination of the movement. The second study investigates more closely how different kinds of expectancy affect the reaching of target points in the squat. In addition, these studies focus on people that struggle with performing such movement.

2. BACKGROUND

Previous work on how technology can be used within physical activity has taken one of two approaches to tackle the problem. The first uses real-time sound feedback to help people correct and optimise their technique during a specific exercise[3, 4, 5, 19, 18, 20]. The aim of these kinds of feedback generally focus on informing people on their deviations from a particular movement path and focus more on the biomechanics of a movement over providing limited motivation or emotional support. The second utilises the promotion of activity through motivational prompts or goal-setting within activity tracking as a way to motivate and increase in one's overall activity levels [9, 10, 11, 12]. While this has shown to be effective in the past, there is also evidence that interventions like this have high rates of abandon and do not provide any information on the quality of a movement but only the amount. In this section, we review the benefits of these two approaches and suggest to extend these two threads of work by investigating how motivational mechanisms can be brought into sonification to enhance motivation on the moment of performing a challenging movement rather than just motivating increasing amount of activity.

2.1. Sonification for Physical Activity

Within sonification, there have been many works in general physical activity in which sonification have been used to inform people of their movement, e.g. [3, 4, 5]. In these works, various sound mappings have been used to inform the individual of their movement. Cesarani et.al used stereo balance in the headphones to represent asymmetry in a swimmers movement, allow them to correct their trajectory[5]. Yang and Hunt used the transformation of sound to provide feedback during the bicep curls on both movement and muscle activation [20]. Hale et.al demonstrate how sound feedback could be used to improve squatting form [18]. Using concurrent feedback on both flexion of the leg and pressure distribution on the foot, they found that participants with the feedback showed more improvement than those without. The prior work demonstrates how movement feedback can be used to improve people's performance of given movements and aid in motor learning. Still, studies have shown that, beyond working as information, sonification of movement milestones may work as a motivational mechanism. For example, in physical rehabilitation, sonification has been used to represent a change in the movement range, for example, the adding layers of music to inform that the range of movement has passed a set threshold [7]. Singh et al [6] went a step further by investigating how the combination of changes in sonification to mark movement milestone and self-calibration of those changes help to increase self-efficacy and sense of safety in people with motor difficulties. Tajadura et al. showed how the altering the feedback heard from one's footsteps can make a person perceive to be lighter led the person to walk faster and perceived the raising of the leg easier [21]. However, there is limited work investigating the motivational mechanisms found within the way sound can be structured as music and how this might be leveraged to motivate and reward people who struggle with physical activity.

2.2. Tracking and Motivational Tools

Biddle's work has shown that it is difficult to understand why it is that people do not adhere to an exercise routine as often only "surface reasons" are given, masking the deeper barriers affecting adherence [8, Ch. 2]. Sonstroem and Morgan highlight the link between exercise and self-esteem [22] in their proposed model for measuring the effect of physical activity interventions on self-esteem. Their model associates the self-perception of physical self-efficacy, competence and acceptance as the contributors to one's self-esteem with regards to physical activity. Additionally, as Shieh et.al show self-efficacy has a direct association with adherence to exercise [23]. Building on the importance of psychological barriers to physical activity, many researchers have proposed technological interventions focus on behaviour change theories to increase the amount of activity people do in their day-to-day lives. As outlined by Sullivan and Lachman, goal setting and rewards are often used in conjunction with activity tracking to motivate increased activity [9]. Goal setting is often used within fitness trackers as a way to motivate people to engage in physical activity [24] and, as shown by Munson and Consolvo [10], can be beneficial within self-monitoring of physical activity. Goal-setting and tracking provide people with a clear target to achieve and measures of their achievement. Additionally, motivation prompts to be more active through text messages or through an automated activity programme have been shown to effectively increase people's activity levels [25, 26]. Furthermore, physical rewards, e.g. a refreshing drink/monetary compensation [11, 12] have shown to impact people's motivation to do physical activity, however, virtual rewards may not have as substantial an impact [10].

While these previous works may demonstrate some efficacy for increasing people's general levels of activity, the focus is on the quantity of activity over the quality. These kinds of intervention, while motivational, do not give any support to facilitate engagement with challenging movements in real-time, which is important not only for optimising technique but for engaging in physical

activity that is beneficial to people well-being and health. Based on the previous success of sonification for real-time feedback to enhance confidence in movement as it is executed and on the importance to understand how motivational mechanisms works, we investigates how musical expectancy added to sonification facilitates the engagement with challenging movement by working not only as a goal-setting and reward mechanisms but also by providing a bottom up embodied desire to move.

3. MUSICAL EXPECTANCY AND MOVEMENT

Music and physical activity have some fundamental connections. Even the way in which music is described in terms of movements and scales moving up and down betrays the embodied way in which we experience music [27, 28]. Moreover, it can be seen how listening to different kinds of music during exercise can have an effect on both affective states and motor performance in physical activity [14, 15]. People have shown an implicit ability to synchronise their movements with music [29] and that certain music can even motivate more exercise/limit the effects of exertion [14, 15]. From this work, it can be seen how powerful a tool musical sound can be in promoting physical activity. People's relationship with music is not inherently linked to formal musical education. Simply through the general everyday exposure, people are able to understand and recognise many aspects of music [30]. Specifically, we look at musical expectancy, which can be described as the way we expect a piece of music to evolve. Whether this expectancy is met or defied creates the relaxations and tensions we feel within music. The work of Bigand demonstrates how harmonic stability, whether the harmony of piece music is complete, and rhythmic stability, whether the piece completes rhythmically, were correctly recognised by both musicians and non-musicians [31]. More recent work by Sears et.al, also shows how the traditional cadences used in classical music can be interpreted by non-musicians [32]. This musical expectancy has been demonstrated to impact people's movement. The work by Komeilipoor et.al [29] showed how musical dissonance impacts people ability to synchronise their movement with an external musical stimulus. It was found that participants were able to better synchronise with a consonant sound (one that fits expectation) than a dissonant sound (one which defies such expectations). Additionally, it was found that the consonant sound improved both form and accuracy. Newbold et.al, demonstrate how aspects of harmonic stability may be built into a movement sonification [16]. Using the stretch forward movement (moving from a neutral standing position and stretching forward to a comfortable target point); it was found that a harmonically stable cadence at the target point promoted the conclusion of the movement and a harmonically unstable sound encouraged additional movement. In addition, it was found that the stable cadences provided a greater sense of reward. However, that work focuses on executing a movement without a clear ending point executed by people that did not have difficulty with it. It is not clear if such mechanism would still work with movements that are supported by clearer interoceptive and exteroceptive feedback mechanism (i.e., proprioceptive and visual perceptual feedback). For example, in the stretch forward exercise (exercise used in Newbold et al.) the movement space has no boundaries and has an open stretching target point, one simply stretches forward into space and could theoretically continue indefinitely. However, the squat down movement offers a closed target that is a 90-degree bending of the knees. Such target position could be said to be signalled by clearer perceptual cues (the flexion of the legs together with the visual feedback with respect to body height rather than depth) and has a definite ending to the movement squat (when the person reaches the squatting position). It is hence to be asked if the expectancy information carried by the sound would still contribute to changes in movement. In addition, Newbold et al. did not investigate the overall effect of expectancy versus sound with no expectancy information on motivation and movement. These are the two questions we address in the following two studies.

In the following, we report two experiments in which we measured both the effect that the sonifications have on the perception of people's own movement and on the quality of the movement itself. Our musical informed sonification leverages people's implicit knowledge of musical endings, i.e. how we expect a piece of music to end. We aim to understand the effect of people's embodied perception of the expectancy of when a music is supposed to end on their movement [27, 16, 29].

4. STUDY 1: MUSICAL EXPECTANCY IN SQUAT SONIFICATION

Our first study investigates if sonification carrying musical expectation, i.e., if or not a piece of music should end, has an impact on motivation to either reach or continue beyond a target position. More specifically, we investigate the use of stable and unstable cadences for defining the target point of the squat movement and compare them to how the squats are performed with either no feedback at all or with feedback that provides no musical expectation, i.e. is purely informative. Based on the previous works showing the effect of musically-informed sonification on physical activity, We hypothesised that:

H1: Sonification will be more motivating than no sound given their pleasurable effect, with the musical sonifications being favoured over white noise.

H2: Unstable cadence will encourage the most additional movement, while the sonification carrying no expectation will have the quickest start of return time from the squat position to standing position;

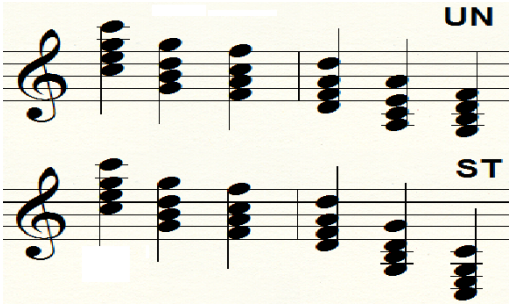


Figure 2: The two chord sequences used to create our two musical conditions, unstable shown top, ending on an imperfect cadence, dominant 7th and stable below, ending on a perfect cadence.

4.1. Materials

The squat movement is tracked using a smartphone strapped to the upper leg, as shown in 1. The on-board gyroscope, 50 FPS,

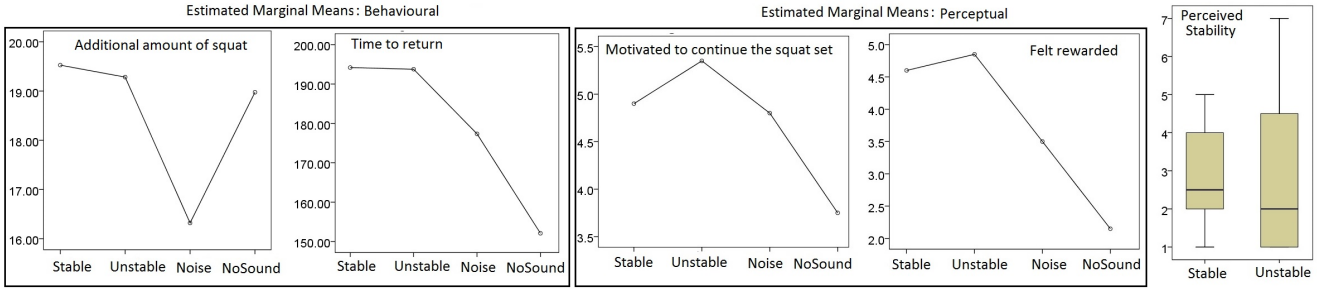


Figure 3: Overview of the finding of study one, both behavioural and perceptual

measured the angle of the leg during the squat. The phone is calibrated between the participant's starting position (i.e., standing in the squat case) and their target squatting position. The calibration consists in dividing the range of movement between the standing position and the target point into 6 movement segments with each segment triggering the next chord, thus playing the full chord sequence as the movement progresses towards the target point and which point the sonification ends (see Figure 2).

Four sonification conditions are defined: two creating expectations of musical endings (Stable, Unstable) and two that create no musical ending expectation (noise as sound and no-sound condition). These are described below:

- Stable (ST) and Unstable (UN) sonifications: As the participant moves through the calibrated space, each chord is played in equidistant intervals with the final either ending in a stable or unstable cadence (perfect or imperfect dominant 7th respectively) as seen in Figure 2.

- Noise or also defined hereafter as non-musical sonification (NM): white-noise was used so as to convey no musical expectation. The white noise sounded during the movement and reaching the target point was signified by the noise stopping. This sonification was used to compare these musical sonifications to one that is purely informative; as such participants would still know when the target point was reached, but there would be no prior expectation of its ending.

- No-sound condition (NS): this condition was considered to compare these sounds to how a squat would be performed unaided.

4.2. Participants

A total of 20 paid participants were recruited for the study (age=20-62 (mean = 26.5), 15 female and 5 male). All participants reported that they did not currently engage in regular physical activity.

4.3. Experimental Design & procedure

The study followed a randomised within-subject design using the four conditions described above. The study measured two behavioural and seven self-reported measures. In terms of behavioural data, the average amount of movement beyond the target point **(Additional Squat)** and the average time taken between the target point and the maximum amount of movement **(Time of Return)** before returning were measured using the smartphone device placed on the upper leg of the participants. For self-reported measures: For all conditions, perceived motivation to continue squatting after the target point (1 for not motivated to 7 for very motivated), motivation to do more squats in the set (1 for not motivated to 7 for very motivated), and perceived reward at the target point (1 for not rewarding to 7 for very rewarding), were taken with 7-point Likert-type response items. Perceived angle of the final squat was reported in degrees, 0 degrees being standing and 90 degrees being the upper leg is parallel to the ground. For the three sound conditions participants were asked how informative the sound was and for the two musical conditions, perceived stability was measured (1 for completely stable 7 completely unstable). The participants were first introduced to the experiment and given a demonstration of the smart-phone application. For each condition participants were asked to set a goal of how many squats they would aim to do. This was done so as to give each participant a practical and safe goal to aim for; however, they were informed that they could stop the set before reaching the goal if the wished and to use the goal as a ballpark number to aim toward. They were instructed to squat down at a steady pace allowing each sound to play and to use the music/sound produced to inform them when they had reached the target point (i.e., the end of the feedback), in the no sound condition they were told to go until they felt they had reached the target point. After each set of squats, a questionnaire was used to collect the self-report measures.

5. RESULTS

Results are summarised in Figure 3 and further details of the descriptive statistics can be found in Figure 4. The behavioural measures were submitted to repeated measures analyses of variance (ANOVA) followed by Bonferroni-corrected pairwise comparisons; 17 observations for each condition were analysed, three participants were removed due to data loss. Self-reported data was analysed with Friedman tests, followed by Bonferroni-corrected Wilcoxon signed rank tests.

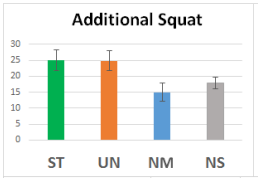
Additional Squat: Significant effects were found for the amount of squat past the target point ($F(3, 48) = 5.36, p = .003, \chi^2 = .251$). Participants moved significantly more in both musical conditions than in the non-musical conditions:

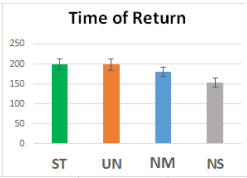
- ST $>$ NM and NS (p=.022 & p=.047).
- UN $>$ NM and NS conditions (p=.014 & p=.029).

However, no differences were found between the two expectancy conditions (i.e., ST vs UN) and between the two non-expectancy conditions (NM vs NS).

Time of return: Significant effects were found across conditions ($F(3, 48) = 10.23, p < .001, \chi^2 = .390$). Significant differences were found:

- ST (stability) $>$ NS (p=.004)





	Stable	Unstable	Non- Musical	No sound
Stability	2.5(1-5)	2(1-7)	n/a	n/a
Informative	4.5(3-7)	5(1-7)	5(1-7)	n/a
Motivation to Squat Deeper	5(1-7)	5.5(1-7)	4(1-7)	2(1-4)
Motivation to do more squats	4(1-7)	4(1-7)	4(1-7)	4(1-6)
Reward	5(1-7)	5.5(1-7)	5(2-7)	4(1-7)
Angle	72.5(40-100)	75(45-95)	80(35-95)	77.5(35-105)

Figure 4: Overview of results from study one, showing the mean additional squat and time of return (SE) for stable, unstable, non-stability and no sound. Bottom shows the median (IQR) of the self-report measures.

- UN (unstable) $>$ both NM (p=.005) and NS ($p < .001$)

This means that people took longer to start the return to a standing position with the cadence sonifications. However, while the stable condition had an effect only on the no-sound condition, the unstable ending did have an effect with respect to noise. At the same time

- NM $>$ NS (p= .018) with people taking longer to start the return in the NM condition.

This suggests that the musicality of the sonification encourage movement in itself beyond the target point in comparison to no-sound or noise. Still, it appears that even with simple noise feedback people show a slower time of return.

Self-reported measures: A series of Friedman tests found significant differences for: perceived motivation to continue (χ^2 (3) = 11.354, p = .010) and perceived reward (χ^2 (3) = 22.288, $p <$ .001). Subsequent Wilcoxon tests (the adjusted significance level was at p = 0.008) showed that:

- motivated to continue the set: UN $>$ NS condition (Z = -2.815, p = .005)
- perceived reward both ST $>$ NS (Z = -3.35, p = .001) UN $>$ NS(Z = -3.42, p = .001)

It should be noted that the two musical conditions (ST and UN) were not found to be significantly different in terms of perceived stability, (Z= -.144, p=.886). These latter results differ from Newbold et al. where participants did perceive differences in stability between these two musical endings [16].

6. DISCUSSION

The results presented above support in part the set hypotheses. The results show some effects in movement behaviour when comparing musical cadences with non-musical feedback/no sound at all. In fact, less movement past the target point and a faster return were found for both the NM and NS conditions. However, no differences were found between the two levels of stability in terms of behaviour, only partially supporting **H1**.

For the self-reported measures, there was also a limited impact of the different stabilities, however, the unstable (UN) cadence was found to motivate participants to do more squats than the no sound condition (NS). In addition, both musical sonification (UN and ST) were found more rewarding than the no-sound condition NS. The shorter movement and quicker return time found in the NM and NS conditions suggest that upon reaching the target point participants began standing back up from the squat. This may suggest that the stopping point (either because of abrupt stopping of the sound or because of proprioceptive feedback) is much clearer in these conditions, leading to an immediate and almost jerky turn around. Conversely, in the musical conditions this ending is perhaps less clear cut, meaning the return takes longer to come about and while this does lead to more movement, it also gives a smoother turning point. The stability for both ST and UN condition was quite low. It is possible that this has invited to continue to move as the sonification built the expectation of sound continuation despite the sound ended. Indeed, people reported being more motivated to continue the movement beyond the target point in the musical condition rather than in the non-expectation.

What is not so clear is why people did not perceive differences in stability levels between the ST and UN conditions. One possibility is that the lack of expectation of sound continuation in the noise condition has led to perceive both musical conditions having a more movement engaging perception. It may also be possible that the fact that the people executed a sequence of squats rather than just one, the music melody was perceived as continuing from one squat to the other leading to a smoother inversion between two consecutive squats. This is supported by the fact that the unstable was perceived as most motivating to continue the set; participants being motivated by the perceived continuation of the music.

In addition, while no differences were identified between conditions in terms of information carried (see figure 4), both musical conditions had a greater impact on the motivation and reward, as hypothesised in **H2**. These results show how these musical sonifications can motivate people during physical activity and how musical sonification can be seen as a reward for completed music. This verifies previous works in sonification findings [16, 6] on the impact of music on motivation during exercise. In the following study, we better investigate the if there is an effect of stability level on movement by exposing participants to only ST and UN conditions, this time for a single squat.

7. STUDY 2: SINGLE SQUAT SONIFICATION

The above study found that while musically informed sonification did impact people's perception of their movement, no differences were found between the two levels of stability. These results are somehow in contradiction with Newbold et al [16] the squat repetitions (versus a single instance of the movement as in Newbold et al) had on the perceived stability or due to the much greater perceived difference between the musical and non-musical sonifications. Therefore, in this study, we explore the effect by exposing participants to only the ST and UN condition and to a single squat scenario. We hypothesised that:

H1: stable cadences will feel more rewarding and increase the sense of achievement, while unstable cadences will trigger more motivation to continue the movement.

H2: Unstable target points will encourage the continuation of the downward movement;

7.1. Materials

The squat movement was tracked in the same way as the above study, see Figure 1. This time only the two musical sonifications were compared. The sonification used a combination of the two

cadence types (stable and unstable) and three lengths (8 chords, 7 chords and 6 chords), used to avoid participants learning the location of the final cadence during the experiment.

7.2. Participants

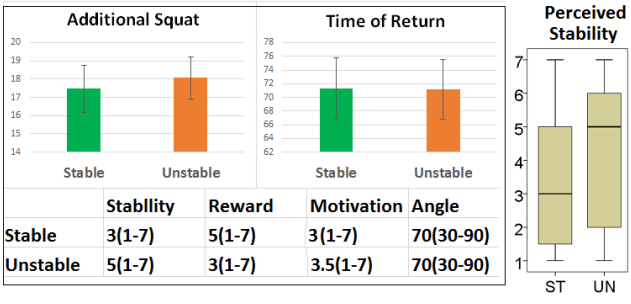
A total of 20 paid participants were recruited for the study (age=19-48 (mean = 28.4), 10 female and 10 male). All participants reported that they did not currently engage in regular physical activity.

7.3. Experimental Design & procedure

The study followed a randomised within-subject design using the six total conditions described above (two cadences (ST and UN) x three lengths). The study measured two behavioural and four self-reported measures. In terms of behavioural data, the additional squat past the target point and the time take to return were measured using the smartphone device. For self-reported measures: Measures of perceived stability (1 for completely stable to 7 completely unstable), perceived motivation to continue the movement (1 for very unmotivated to 7 very motivated) and perceived reward (1 for not rewarding to 7 for very rewarding) at the target point were taken with 7-point Likert-type response items. The perceived angle of the squat was reported in degrees, 0 degrees being standing and 90 degrees being the upper leg is parallel to the ground. The participants were first introduced to the experiment and given a demonstration of the smart-phone application. They were instructed to squat down at a steady pace allowing each chord to sound and to use the music produced to inform them when they had reached the target point (i.e., the end of the feedback). After each squat a questionnaire was used to collect the self-report measures.

8. RESULTS

Descriptive statistics for all measures are summarised in Figure 5. The behavioural measures were submitted to repeated measures analyses of variance (ANOVA) followed by Bonferroni-corrected pairwise comparisons;. Self-reported data were averaged across lengths and analysed with Wilcoxon signed rank tests.



	Stablity	Reward	Motivation	Angle
Stable	3(1-7)	5(1-7)	3(1-7)	70(30-90)
Unstable	5(1-7)	3(1-7)	3.5(1-7)	70(30-90)

Figure 5: Overview of results from study two, showing the mean additional squat and time of return (SE), top and the median (IQR) of the self-report measures.

Additional Squat: No significant effects were found for the amount of squat past the target point across the two stability's. However the length condition did have a significant impact of the amount of additional squat ($F(2, 38) = 8.14, p = .001, \chi^2 = .300$). With the pairwise comparisons showing a significant difference between the long and short length (Z = .842, $p < .001$) and the middle and short length (Z = .723, p =.001).

Time of return: No significant effects were found for the amount of squat past the target point across the two stability levels. However the length condition did have a significant impact of the time of return ($F(2, 38) = 4.74, p = .015, \chi^2 = .200$). With the pairwise comparisons showing a significant difference between the long and short length (Z = 21.33, $p < .034$) and the middle and short length (Z = .723, p =.001).

Self-reported measures: Wilcoxon tests showed that:

- Perceived Stability ST > UN (Z = -2.65, p = .014).
- Perceived motivation to squat further UN > ST endings (Z = -2.45, p = .014).
- Perceived depth of the Squat ST >UN (Z = - 2.36, p = .018).

9. DISCUSSION

These two studies demonstrate how musically informed sonifications may be used to support general physical activity and specifically their effect on the squat movement. While previous works have shown the efficacy of sonification in improving the squat movement [18], these musical sonifications aim to concurrently announce progress and provide motivation. The results of our second study did not demonstrate how musical completion at the end of the movement can create a sense of reward, as utilised previously in work focused on physical rehabilitation [33, 16]. However, it was found that participants perceived that they had squatted deeper in stable conditions, even though no such difference was found in the actual movement. This may be due to the sense of completion leaving participants feeling they have moved further may come from the expectancy they feel in the cadences, the unstable being incomplete making people feel like they had completed less movement.

Additionally, it demonstrates how unstable endings provide a motivation to continue which we can link to previous works that show how the response time to unexpected endings is longer than complete ones [34]. These results, support **H1** and suggest that the use of this kind of feedback could improve self-efficacy in people engaging in physical activity; as well as acting as a motivational tool during the movement. While in the first experiment, this effect was lessened, likely due to the lack of perceived difference in the stability, both the musical sonifications were still perceived as more rewarding than both the control of no sound and to the non-musical feedback. The unstable ending then also motivated people to continue the set of squats more, perhaps as the incomplete ending of the sonification made the future repetitions more natural.

However, whereas Newbold et al [16] found that stable cadences promoted the conclusion of a movement, while unstable cadences encouraged additional movement in the stretch forward exercise, we do not find evidence to support **H2** as the same was not true in squat down exercise in this or the previous study. This could be related to the use of additional perceptual cues used by participants. The may also explain why significant differences where found for different length stimulus, as participants would move to the same point regardless. As the stretch forward is a more open-ended movement, the cadences stability has a bigger impact on movement.

As noted above, there was a difference in the way participants perceived the sonifications when doing a single squat versus the full set. Most notably it seems the difference in the two stabilities was not perceived by participants during the sets of squats. This finding disagrees with previous research both on how people perceive harmonic stability [31, 32] and how it is perceived during exercise [16]. However, in the single squat study, the difference in stability was perceived. While may come from the impact of the non-musical stimulus in study one affecting participants perception of the musical sonifications, it could be that the repetition of the movement impacted how the stability of the music was heard; musically this is understandable as the phrase would no longer be heard in isolation, as shown in Figure 2, but would be heard as a single longer piece of music comprises the phrases in sequence. This may also what cause the unstable conditions in study one to have the greatest impact on motivation to continue the set, due to the feeling of musical continuation. This may also account for the changes in perception between the two studies, as in the first experiment the unstable ending was heard as a continuous piece of music but the stable endings meant that where one would expect the sound to be driven forward, there was an unnatural stopping point.

These results of these studies also point toward new questions about how these sonifications may be applied to certain kinds of movements and how the repetition of movements may impact how they are perceived. Future work will include investigating how this kind of sonification can be designed for repetition based exercise. For example, rather than using static sonifications as presented here, sonifications could be designed to evolve and develop throughout a set of repetitions to fully take advantage of people's embodied interaction with these sonifications. Further investigation of categories of movement targets (i.e. open/closed) may enable greater exploitation of musical strategies to support physical activity. Additionally, the impact of musically informed sonifications on people's exercise patterns over the long-term would be interesting to study, particularly in terms of increasing people's self-efficacy over time and their ability to maintain a routine.

10. CONCLUSION

This paper presents a musically-informed sonification that uses musical expectancy to support people who struggle with physical activity. We present two experiments that investigate how musical expectancy impacts people who struggle with physical activity during the squat down exercise. This work demonstrates how musically-informed sonification can support motivational needs of people struggling with physical activity. In additional we show that by using musical expectation people feel they have achieved more movement (in the stable case). While in the unstable case people are motivated to move more and motivated to do more repetitions. However, changes in movement behaviour are less pronounced in the squat down exercise than exercises where perceptual cues at the target are weaker, as found by Newbold et.al [16].

In conclusion, this paper presented a musically-informed sonification of movement in the context of general exercise and the results of studies that investigate how it can be applied to the squat down exercise. This work shows both how the use of musical expectancy can be used to provide in-the-moment motivation and how the impact of perceptual cues within a movement and repetitions should be considered when designing this kind of sonification. The use of musical expectancy within sonification demonstrates that we can combine the power of sound as a feedback mechanism with motivational aspects of music to support people who struggle with physical activity.

11. ACKNOWLEDGMENT

Supported by EPSRC EP/H017178/1 'Emo& Pain' and by an EPSRC DTG.

12. REFERENCES

[1] S. Allender, C. Foster, P. Scarborough, and M. Rayner, "The burden of physical activity-related ill health in the UK." *Journal of epidemiology and community health*, vol. 61, no. 4, pp. 344–8, Apr. 2007. [Online]. Available: http://jech.bmj.com/content/61/4/344.full

[2] S. DellaVigna and U. Malmendier, "Paying not to go to the gym," *The American Economic Review*, 2006. [Online]. Available: http://www.jstor.org/stable/30034067

[3] K. M. Smith and D. Claveau, "The Sonification and Learning of Human Motion," *20th International Conference on Auditory Display (ICAD 2014)*, 2014. [Online]. Available: http://smartech.gatech.edu/handle/1853/52049

[4] N. Schaffert, K. Mattes, and A. O. Effenberg, "The sound of rowing stroke cycles as acoustic feedback," in *Proceedings of the 17th International Conference on Auditory Display (ICAD 2011)*, 2011.

[5] D. Cesarini, T. Hermann, and B. E. Ungerechts, "An interactive sonification system for swimming evaluated by users," in *SoniHED- Conference on Sonification of Health and Enviromental Data*, 2014.

[6] A. Singh, S. Piana, D. Pollarolo, G. Volpe, G. Varni, A. Tajadura-Jimnez, A. C. Williams, A. Camurri, and N. Bianchi-Berthouze, "Go-with-the-flow: Tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain," *HumanComputer Interaction*, vol. 31, no. 3-4, pp. 335–383, 2016. [Online]. Available: http://dx.doi.org/10.1080/07370024.2015.1085310

[7] K. Vogt, D. Pirrò, I. Kobenz, R. Höldrich, and G. Eckel, "PhysioSonic - Evaluated movement sonification as auditory feedback in physiotherapy," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5954 LNCS, 2010, pp. 103–120.

[8] S. Biddle and N. Mutrie, *Psychology of physical activity: Determinants, well-being and interventions*, 2007.

[9] A. N. Sullivan and M. E. Lachman, "Behavior change with fitness technology in sedentary adults: A review of the evidence for increasing physical activity," *Frontiers in Public Health*, vol. 4, p. 289, 2017. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fpubh.2016.00289

[10] S. A. Munson and S. Consolvo, "Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity," in *Pervasive computing technologies for healthcare (PervasiveHealth), 2012 6th international conference on*. IEEE, 2012, pp. 25–32.

[11] R. A. Khot, J. Lee, D. Aggarwal, L. Hjorth, and F. Mueller, "Tastybeats: Designing palatable representations of physical activity," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 2933–2942.

[12] M. Patel, D. Asch, R. Rosin, and et al, "Framing financial incentives to increase physical activity among overweight and obese adults: A randomized, controlled trial," *Annals of Internal Medicine*, vol. 164, no. 6, pp. 385–394, 2016. [Online]. Available: +http://dx.doi.org/10.7326/M15-1635

[13] L. B. Meyer, *Emotion and meaning in music*. University of chicago Press, 2008.

[14] C. Karageorghis and P. Terry, "The psychophysical effects of music in sport and exercise: A review," *Journal of Sport Behavior*, 1997.

[15] H. Mohammadzadeh, B. Tartibiyan, and A. Ahmadi, "The effects of music on the perceived exertion rate and performance of trained and untrained individuals during progressive exercise," vol. 6, no. 1, jan 2008.

[16] J. W. Newbold, N. Bianchi-Berthouze, N. E. Gold, A. Tajadura-Jiménez, and A. C. Williams, "Musically informed sonification for chronic pain rehabilitation: Facilitating progress & avoiding over-doing," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 5698–5703. [Online]. Available: http://doi.acm.org/10.1145/2858036.2858302

[17] G. D. Myer, A. M. Kushner, J. L. Brent, B. J. Schoenfeld, J. Hugentobler, R. S. Lloyd, A. Vermeil, D. A. Chu, J. Harbin, and S. M. McGill, "The back squat: A proposed assessment of functional deficits and technical factors that limit performance," *Strength and conditioning journal*, vol. 36, no. 6, p. 4, 2014.

[18] R. Hale, J. Hausselle, and R. V. Gonzalez, "Impact of error sonifcation auditory feedback for neromuscular training across two degrees of freedom," in *Proceedings of the Annual American Society of Biomechanics, 2015*.

[19] K. Oakes, K. A. Siek, and H. MacLeod, "Muscle-Memory: identifying the scope of wearable technology in high intensity exercise communities," pp. 193–200, May 2015. [Online]. Available: http://dl.acm.org/citation.cfm?id=2826165.2826193

[20] J. Yang and A. Hunt, "Real-time auditory feedback of arm movement and EMG in biceps curl training to enhance the quality ," in *SoniHED- Conference on Sonification of Health and Enviromental Data*, York, U.K, 2014.

[21] A. Tajadura-Jiménez, M. Basia, O. Deroy, M. Fairhurst, N. Marquardt, and N. Bianchi-Berthouze, "As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait," pp. 2943–2952, 2015.

[22] R. J. Sonstroem and W. P. Morgan, "Exercise and self-esteem: rationale and model." *Medicine and science in sports and exercise*, vol. 21, no. 3, pp. 329–37, 1989. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2659918

[23] C. Shieh, M. T. Weaver, K. M. Hanna, K. Newsome, and M. Mogos, "Association of self-efficacy and self-regulation with nutrition and exercise behaviors in a community sample of adults," *Journal of community health nursing*, vol. 32, no. 4, pp. 199–211, 2015.

[24] K. Mercer, M. Li, L. Giangregorio, C. Burns, and K. Grindrod, "Behavior change techniques present in wearable activity trackers: a critical analysis," *JMIR mHealth and uHealth*, vol. 4, no. 2, 2016.

[25] N. Notthoff and L. L. Carstensen, "Promoting walking in older adults: Perceived neighborhood walkability influences the effectiveness of motivational messages," *Journal of Health Psychology*, 2014. [Online]. Available: http://dx.doi.org/10.1177/1359105315616470

[26] R. Hurling, M. Catt, M. De Boni, B. Fairley, T. Hurst, P. Murray, A. Richardson, and J. Sodhi, "Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial," *Journal of medical Internet research*, vol. 9, no. 2, p. e7, 2007.

[27] M. Leman, *Embodied music cognition and mediation technology*. Mit Press, 2008.

[28] L. M. Zbikowski, "Music, language, and multimodal metaphor," *Multimodal metaphor*, pp. 359–381, 2009.

[29] N. Komeilipoor, M. W. M. Rodger, C. M. Craig, and P. Cesari, "(dis-)harmony in movement: effects of musical dissonance on movement timing and form," *Experimental Brain Research*, vol. 233, no. 5, pp. 1585–1595, 2015. [Online]. Available: http://dx.doi.org/10.1007/s00221-015-4233-9

[30] P. Polotti, D. Rocchesso, and D. R. Editors, *Sound to Sense , Sense to Sound A State of the Art in Sound and Music Computing*, 2008. [Online]. Available: http://smcnetwork.org/public/S2S2BOOK1.pdf

[31] E. Bigand, "Perceiving musical stability: The effect of tonal structure, rhythm, and musical expertise." *... of Experimental Psychology: Human Perception and ...*, 1997. [Online]. Available: http://psycnet.apa.org/journals/xhp/23/3/808/

[32] D. Sears, W. E. Caplin, and S. McAdams, "Perceiving the classical cadence," *Music Perception: An Interdisciplinary Journal*, vol. 31, no. 5, pp. 397–417, 2014.

[33] R. I. Wallis, T. Ingalls, T. Rikakis, L. Olsen, Y. Chen, W. Xu, and H. Sundaram, "Real-time sonification movement for an immersive stroke rehabilitation environment," in *Proceedings of the 13th International Conference on Auditory Display (ICAD 2007)*, 2007, pp. 497–503. [Online]. Available: http://www.icad.org/node/2460

[34] E. G. Schellenberg, E. Bigand, B. Poulin-Charronnat, C. Garnier, and C. Stevens, "Children's implicit knowledge of harmony in Western music." *Developmental science*, vol. 8, no. 6, pp. 551–66, Nov. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16246247

EMPIRICALLY INFORMED SOUND SYNTHESIS APPLICATION FOR ENHANCING THE PERCEPTION OF EXPRESSIVE ROBOTIC MOVEMENT

Jon Bellona

University of Virginia
Department of Music
Charlottesville, VA, USA
jpbellona@virginia.edu

Lin Bai

University of Virginia
Electrical and Computer Engineering
Charlottesville, VA, USA
lb7ss@virginia.edu

Luke Dahl

University of Virginia
Department of Music
Charlottesville, VA, USA
lukedahl@virginia.edu

Amy LaViers

University of Illinois
Mechanical Science and Engineering
Champaign-Urbana, IL, USA
alaviers@illinois.edu

ABSTRACT

Since people often communicate internal states and intentions through movement, robots can better interact with humans if they too can modify their movements to communicate changing state. These movements, which may be seen as supplementary to those required for workspace tasks, may be termed "expressive." However, robot hardware, which cannot recreate the same range of dynamics as human limbs, often limit expressive capacity. One solution is to augment expressive robotic movement with expressive sound.

To that end, this paper presents an application for synthesizing sounds that match various movement qualities. Its design is based on an empirical study analyzing sound and movement qualities, where movement qualities are parametrized according to Laban's Effort System. Our results suggests a number of correspondences between movement qualities and sound qualities. These correspondences are presented here and discussed within the context of designing movement-quality-to-sound-quality mappings in our sound synthesis application. This application will be used in future work testing user perceptions of expressive movements with synchronous sounds.

1. INTRODUCTION

Humans often communicate intentions and affective states through qualitative aspects of their body movement. For example, the same hand movement can elicit two different responses – warmth, as when one points to a friend on the street, or aggression, as when one points out the accused in a police lineup. The difference is in the context and *how* the gesture is executed.

https://doi.org/10.21785/icad2017.049

Robots in human-facing roles could more effectively interact with humans if they too could express various intentions through their movement. Ongoing research is being conducted with the goal of endowing robots with the ability to express different qualities in their movement [1, 2]. However, the physical limitations of specific robotic platforms can reduce a robot's ability to sufficiently express differences in movement qualities [3].

Humans also communicate intentions and emotions sonically through non-verbal aspects of their vocalizations. For example, an angry person might modulate the loudness and resonance of their voice, which mirrors shifts in body language and emotion. Thus, by endowing robots with expressive sound we may improve people's perception of varying qualities in robotic movement. The question then becomes, how does one design sounds to accompany expressive movements with perceptual cohesion?

In this paper we present a quantitative, empirically-based sound design approach and a synthesis application for generating sounds to accompany various movement types. We summarize a study we conducted [4], whose results suggest a number of correspondences between movement qualities and sound qualities. And we describe how these correspondences are used to control sound synthesis parameters in our application.

In our study we asked experienced musicians to vocalize sounds to animations of a simple movement. Different versions of this movement were generated by varying the Effort Factors proposed by movement theorist Rudolf Laban (see Section 2.1). The musicians' vocalizations were recorded and then analyzed by manually applying qualitative labels to each sound, and by performing a quantitative signal analysis using techniques from MIRtoolbox [5]. Our goal was to discern general trends in how sounds differ as movement quality varies.

We applied the study results to a sound synthesis application in which we dynamically modify sound quality parameters according to changes in movement quality. We control our sound synthesis with the same movement quality parameters that are used to generate movement trajectories for robotic movement (as described in [3]).

2. BACKGROUND

A number of studies have been conducted on movement sonification. In music, studies of body movements of acoustic musicians by [6] show correlations of qualitative movement related to structural components (e.g. tempo) within musical works. Another study correlates expressive movement to musical phrasing by tracing movement shapes of nine pianists [7]. Our work moves the other way in taking movement qualities and mapping these onto their respective parameters of sound qualities in a sound synthesis application.

Movement sonification studies in athletic tasks demonstrate acute human perception of changes in timing with movement sonification [8] [9]. Instead of sonifying repetitive sequenced patterns of movements, our sound design focuses on generating sounds based upon short, singular movements. Effenberg et al. [10] used a direct mapping of "kinematic and dynamic motion parameters to electronic sounds" to help improve motor learning in sports. While we also integrate multi-modal perception into our work, we are interested in designing sounds that correspond to basic movement qualities instead of designing for specific sequenced movement structures used in specific tasks.

Similarly, [11] and [12] measured kinematic features of movement and transformed these features into sound. And [13] presents a real-time sonification framework for all common MIDI environments based upon acceleration and orientation data from inertial sensors. Our sound synthesis integrates Laban Effort Factors, which describe how a movement is performed, with the aim to improve the perception of expressive robotic movement.

Generating robotic movement by mapping the components of Laban's Effort system to weighted parameters in an optimal control problem builds on [14]. Our sound synthesis application parameters leverage sound qualities corresponding to movement qualities from an analysis of vocalizations of movement by [4]. These movement to sound quality correspondences will be discussed further in Section 3. The approach toward generating sounds for preexisting movements is similar to work in [15] in that they use Laban Effort Factors to organize movement; however, their participants train a system with movement data in order to generate sound from dancers' movements in real time. Our work aims to generate sound from movement quality parameters.

Lastly, we acknowledge a rich history of sound mappings for sonification purposes [16]. Sound design is a skilled art. Our goal isn't *musical* sound design, but rather to communicate intent and affective state with sound along with movement based upon their perceptual links. Rather than using aesthetics to drive mapping choices as seen in [17] and [18], our mapping choices are instead based upon quantitative signal analyses and qualitative results (see Sections 3.1 and 3.2).

2.1. Describing Movement Quality

Our synthesis framework, as well as the algorithms we use for generating robotic movement [3], are controlled by parameters based on the Effort system defined by Rudolf Laban [19, 20] to qualitatively specify the ways in which a movement may vary. In this system, movements are characterized using four Effort Factors: Space Effort, Time Effort, Weight Effort, and Flow Effort. We capitalize these terms to avoid possible confusion with other notions of these terms.

Space Effort describes the attitude toward the environment of a movement. A movement can be Direct (as in throwing a boxing jab), or Indirect (as in shaking out a rock-filled boot). Time Effort describes the attitude towards initiation and completion of a movement. A movement can be Sudden (as in pulling your hand away from a hot stove) or Sustained (as in pushing against a grand piano). Weight Effort describes the attitude towards the mover's mass. A movement may be Strong (as in sprinting a 100 meter dash) or Light (as in touching a baby's cheek). Flow Effort describes the progression of a series of movements. We do not use Flow Effort in our sound application, since the movements analyzed for sound synthesis may all be considered singular movements.

Table 1: Laban's Eight Basic Effort Actions

Movement	Time	Space	Weight
Gliding	Sustained	Direct	Light
Pressing	Sustained	Direct	Strong
Floating	Sustained	Indirect	Light
Wringing	Sustained	Indirect	Strong
Dabbing	Sudden	Direct	Light
Thrusting	Sudden	Direct	Strong
Flicking	Sudden	Indirect	Light
Slashing	Sudden	Indirect	Strong

When paired in all combinations, the three Effort Factors that we employ – Space Effort, Weight Effort, and Time Effort – form Laban's eight Basic Effort Actions (BEAs), as shown in Table 1. The BEAs are Dabbing, Flicking, Floating, Gliding, Pressing, Slashing, Thrusting, and Wringing. These eight actions constitute a set of basic movements whose qualities people can understand by drawing on their own experiences [20].

3. A STUDY ON SOUND MOVEMENT CORRESPONDENCES

This paper documents an application to generate sounds based upon correspondences between movement and sound, so that we may further study the perception of endowing expressive robotic movement with expressively coherent sound. In order to search for these correspondences we conducted an initial study, whose details are presented in [4], and which we summarize here.

We created animations of a stick figure performing a simple movement with the qualities of each of the eight BEAs. The figure moves from a pose where the hands are near the center line, to a pose where the arms are extended to the sides, and then returns to the first pose. Each movement was four seconds long. The qualitative variations in the movement trajectories are created by changing three parameters of a control algorithm described in [3] and [14]. These parameters correspond to the Laban Effort Factors of Space, Weight, and Time.

We presented these animations to seven musicians (graduate students and professional musicians, all with significant improvisation experience), along with the BEA label, and asked them to vocalize a sound that matched the qualities of the movement and label. We recorded these sounds and subjected them to the analyses described in Section 3.1. The BEA labels were necessary because the animations, though generated with different movement qualities, were not different enough for the participants to be able

to easily distinguish the differences. Indeed, this is part of the motivation for this work.

3.1. Quantitative and Qualitative Sound Analysis

We performed two analyses of the recorded vocalizations. In the first, we manually applied qualitative labels to each recording, and in the second we performed signal analysis in order to quantify various sonic qualities.

In the first analysis, each of the four authors listened to and manually applied qualitative labels to each of the 56 recordings (7 musicians $\times$ 8 BEAs). We chose a set of labels for the attributes of pitch, loudness, and timbre, as well as for the shape of how pitch, loudness, and timbre varied over each sound's duration. The labels we used can be seen in Figure 1. For each attribute, a listener was allowed to apply only one label. In order to look for meaningful sonic differences between movement qualities, we reorganized the label data according to each Effort Factor (Space, Weight, Time), and created histograms to compare the two values of each Factor. For example, in Figure 1 we can compare Strong and Light Weight Efforts, and we see that the amplitude of sounds for Strong Effort movements are more often labelled 'medium' and 'loud' than are sounds for Light Effort movements.

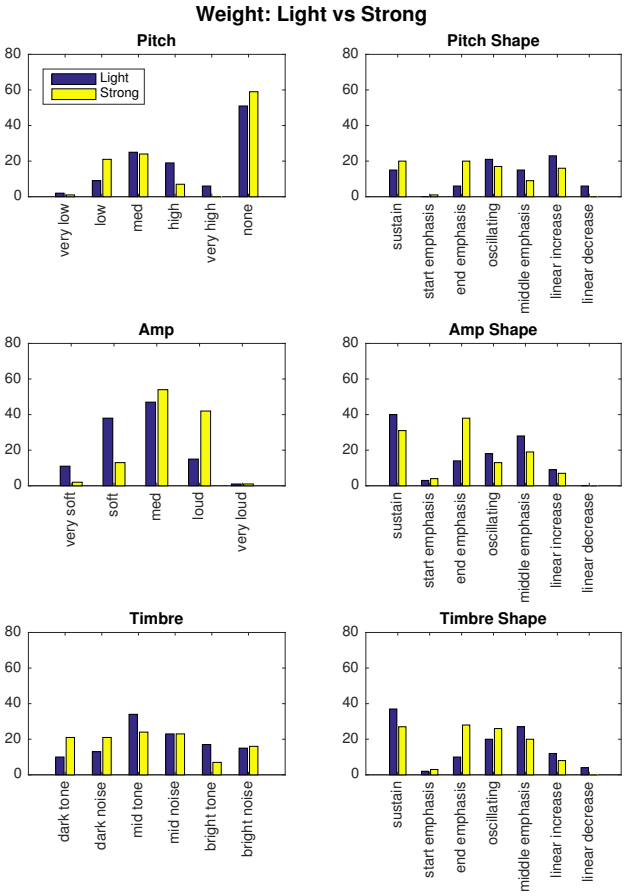


Figure 1: Qualitative label counts for the Effort Factor Weight comparing Strong vs. Light

In the quantitative analysis, we used the MIRtoolbox [5] to extract the following audio features for each recording: amplitude envelope, spectral brightness, spectral centroid, spectral rolloff, spectral flatness, and zero-crossing rate. We calculated the mean value of each feature across each recording, which we call the 'recording mean'. To account for differences in vocal range or performance style, for each participant we subtracted out the mean of each feature taken across all recording means for that participant. We then reorganized the data by Effort Factors, which allows to conduct T-tests to determine if a features varies significantly between the two values of a given Effort Factor. For example, Figure 2 shows that the mean spectral rolloff for movements whose Time Effort is Sudden is higher than for movements whose Time Effort is Sustained, and a T-test confirms this ($t(54) = 2.74995, p = 0.0081$).

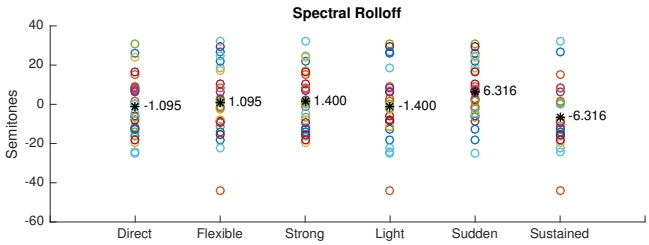


Figure 2: Spectral Rolloff for the Space, Weight, and Time Effort Factors.

3.2. Sound Movement Correspondence Findings

Detailed results of our study are described in [4]. Here we summarize the findings, which are used to inform the design of our sound synthesis application. P-values for significant findings from the quantitative analysis are shown in Table 2.

Here we describe the findings for each Effort Factor, and indicate whether they came from the qualitative (*qual*) or quantitative (*quant*) analysis.

Time Effort:

- Sudden movements are associated with brighter sounds, whereas sounds for Sustained movements tend to have darker timbres (*quant* and *qual*).
- Sudden movements are associated with noisier sounds, whereas sounds for Sustained movements are more pitched (*quant* and *qual*).
- Sounds for Sudden movements tend to have moments of strong emphasis both in amplitude and timbre, whereas sounds for Sustained movements tend to have smooth amplitude envelopes and smoothly varying timbre (*quant* and *qual*).
- Sounds for Sudden movements tended to be louder than for Sustained movements (*qual*).

Weight Effort:

- Strong movements are associated with sounds that are louder and have higher peak amplitudes, whereas sounds for Light movements are quieter and have smaller peak amplitudes (*quant* and *qual*).
- Strong movements are associated with darker sounds, whereas Light movements are associated with brighter sounds (*qual*).

- Sounds for Strong movements tend to have stronger end emphasis, whereas sounds for Light movements tend to have middle emphasis or were sustained (*qual*).

Interestingly, we did not find any notable differences when comparing Direct and Flexible Space Effort. This was true for both the qualitative and quantitative analyses. We hypothesize that this may be true because the concept of space does not map very precisely onto the qualities of sounds we studied, and a full discussion is available in [4].

4. SYNTHESIS APPLICATION

We created a sound synthesis application based upon correspondences between movement and sound (see Section 3). Sound quality parameters directly correlate to movement quality parameters from the movement trajectory function of [3], so that single robotic movement trajectories may be sonified and recorded; although, the software was designed with the potential to translate multiple movement trajectories into sound.

We built our synthesis application with Max/MSP™ software. The application was primarily built using native Max 6 objects, but utilizes scaling functions developed by [21]. We chose electronic synthesis over a sample-based approach in order to retain the most parametric control over our designed sounds. Nearly all audio components are informed by our previous study [4], and parameter ranges used to match the results are based in part by active listening and iteration, as well as psychoacoustic phenomena. Next, we will describe the audio circuit before discussing each Effort Factor parameter in more detail.

4.1. Audio Circuit (Signal Flow)

The sonification uses a combination of additive and subtractive synthesis techniques. We start by generating pitched sounds with three summed oscillators. A rectangle wave sounds the fundamental frequency, and two sawtooth waves sit an octave above (with one of the sawtooth waves being slightly detuned). The summed oscillators move through a Moog™ ladder resonant filter, and movement qualities control the fundamental frequency of the sound, resonant frequencies of the filter, and the sound's amplitude envelope (ADSR) (described in section 4.3).

Separately, a white noise generator feeds a state-variable filter with a center frequency of 4.1 kHz. Both pitch and noise sounds have an ADSR envelope and duration controls to regulate emphasis peaks of amplitude, and each have separate gain controls. After summing noise and pitch signals, the combined signal splits. One signal is kept 'dry' (no effect) and the other signal feeds a detuned delay (chorus effect) that also spreads in the stereo field. Effected and dry signals are combined in the left and right channels, which feed a low-pass biquad filter. A final gain control determines overall amplitude of the sound. For the full schematic, see Figure 3.

4.2. User Parameters

The application interface provides the user with several user parameters, chiefly three main inputs, in the form of Laban Effort Factors: Weight, Time, and Space. These parameter values have a range between 0-1000, with extremes of each range matching movement qualities labels from Table 1. The value ranges in the synthesis application match movement quality parameter values used for generating movement trajectories [3].

Because movement trajectories depend on robotic platforms to carry out generated movements, physical differences between platforms may alter a movement's duration. In addition, the shape of the generated movement does not indicate where in the movement the emphasis will occur, only that an emphasis will occur. For example, the impulse of a Sudden movement may occur at the beginning or end of the movement, but both movements may be deemed as Sudden.

Two additional user parameters affect overall duration and time location of peak amplitude emphasis. These two envelope parameters are set manually in order to match singularly generated movement trajectories. For multiple movement recordings, these two user parameters may need to be automatically updated in order to dynamically respond to trajectory inputs. For now, the sound synthesis application has been used to record single movement sounds, and these sounds will be used for a forthcoming study measuring the perception of our quantitative sound design and movement trajectories. The next section will discuss movement qualities and their associative sound controls.

4.3. Laban Effort Factor Sound Controls

4.3.1. Time

Time Effort values affect amplitude peak and gains for pitched and non-pitched signals as well as overall brightness. These controls correlate to quantitative results of sound qualities shown in Table 2. For amplitude peaks, Sudden Time Effort generates stronger, or emphasized, amplitude peaks and Sustained Effort generates a smooth amplitude envelope curve, void of any peaks. Time Effort simultaneously controls gains for both pitched and non-pitched signals. Sudden sounds tend to be noisier, so that Sudden Effort emphasizes noise signals, with white-noise gain set to 0dBFS and electronically pitched tone signal gain set to -12dBFS. Sustained Effort inversely emphasizes pitched signals, with pitched sound signal gain at 0dBFS and white-noise sound signal gain at -12dBFS.

Time Effort also affects overall brightness by controlling a cut-off frequency of a low-pass bi-quad filter. Sudden Effort maintains the high-end of sounds, whereas Sustained Effort rolls off higher frequencies to produce darker sounds. The large roll-off range of 1kHz – 10kHz was chosen in order to be *perceptible enough* for most listeners.

4.3.2. Weight

Weight Effort controls the peak amplitude of pitched and non-pitched signals, oscillator frequencies and resonant filter frequency of pitched signals, as well as a fine adjustment control of overall amplitude. For peak amplitude, Weight Effort adds additional peak emphasis to signal envelopes, and the peak envelope parameter stands on the quantitative correlation shown in Table 2. Thus, Weight Effort combines with Time Effort in controlling the envelope shape. Time Effort controls a 15dB range for peak emphasis with Weight Effort adding an additional 3dB range onto the peak amplitude.

We found qualitative correlations between Weight Effort and timbre as well as between Weight Effort and pitch [4]. For example, Strong Weight Effort sounds tended to contain more dark tone and dark noise labels, whereas Light Weight Effort sounds had more mid and bright tone labels. While the two Weight Effort qualities share pitch in our correspondence findings (section

Table 2: P values for significant findings comparing Effort Factors to sound qualities

Feature	Effort Factor	P value	Result Summary
Amp. Envelope Entropy	Time	< 0.001	Sudden sounds tend to contain strong peaks
Spectral Flux Entropy	Time	< 0.001	Sudden sounds tend to contain peaks of intense change
Brightness	Time	< 0.01	Sudden sounds tend to be brighter
Spectral Centroid (Log Hz)	Time	< 0.01	Sudden sounds tend to contain higher frequencies
Spectral Flatness	Time	< 0.01	Sudden sounds tend to be noisier
Spectral Roll-off (Log Hz)	Time	< 0.01	Sudden sounds tend to contain more high frequencies
Zero-crossing Rate (Log Hz)	Time	0.0198	Sudden sounds tend to be noisier
Envelope Peak (dB)	Weight	0.0261	Strong sounds tend to contain louder peak values

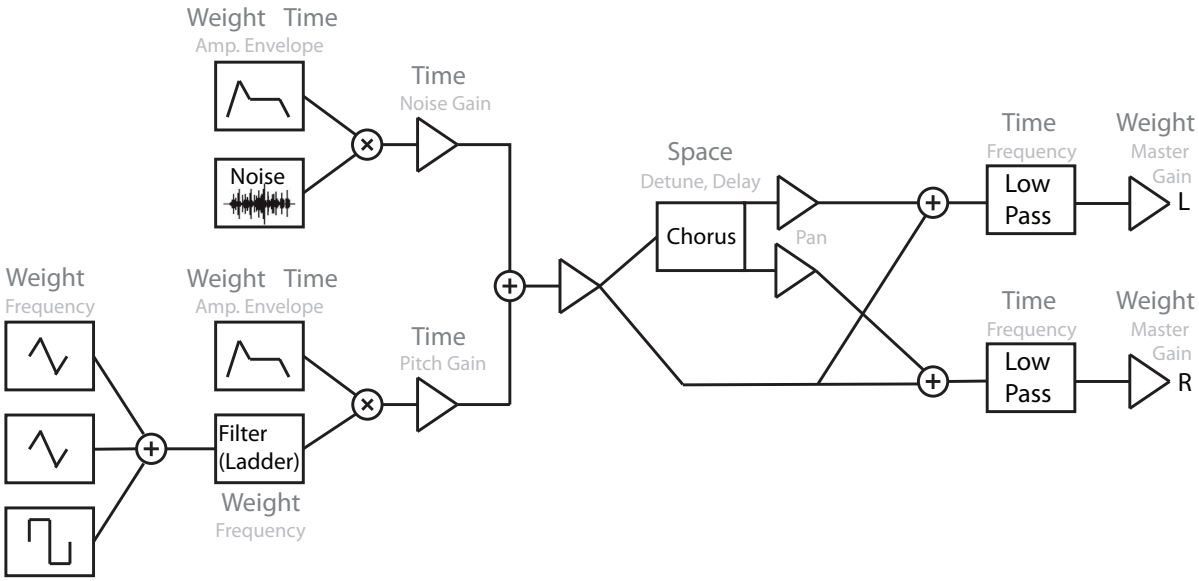


Figure 3: Schematic diagram for movement sonification with Laban Effort Factor (Space, Weight, Time) indicators for parameter control.

3.2), differences exist between pitch and brightness. For example, there is a qualitative difference between a bright tone and a high-pitched tone. Brightness may correspond to noise or pitched signals, whereas pitched tones refer mainly to tonal signals (repeating waveforms). In order to differentiate between pitch and brightness (where brightness is already controlled by Time Effort), oscillator frequency and resonant filter frequency are controlled by Weight Effort. Specifically, Weight Effort affects the sound's fundamental frequency, from 130.81Hz to 523.25Hz, representing a two-octave range between Strong and Light Weight sounds. The two-octaves includes portions of all vocal ranges within its scope [22]. Due to the difference in perceived loudness at high and low frequencies as outlined by Fletcher Munson curves [23], Weight Effort slightly adjusts the master gain in order to account for these differences. We use the nominal mixing level of 80dB to calculate the difference between fundamental frequency of pitched sounds and the state-variable filter of noise sounds, such that for Strong Weight Effort sounds, the master gain is equivalent to 0dBFS and for Light Weight Effort sounds, the master gain is -6dBFS.

Since timbre labels differ for pitch in our correspondence findings, Weight Effort also controls a resonant filter frequency, a Moog™ ladder filter, where the resonant filter alters the timbre of a pitched tone. The range of the resonant frequency is from Eb4 (311.13 hz) to G#7 (3322.4 hz), based upon formant frequency averages by [24] and [25]. The use of a rectangular wave and two sawtooth waves in the generation of the pitched sound creates enough spectral content for the Moog™ ladder filter to effectively output identifiably different resonant frequencies.

4.3.3. Space

For Space Effort, we found no substantial quantitative nor qualitative correlations between movement and sound qualities. However, in order to test possible perceptual indicators of Space Effort in a future study, we chose to include parametric support for Space Effort based upon 'intuitive' mappings. Spatial placement of sound is a creative dimension of music composition, and spatial metaphors are often used in describing music and our experience of listening to music [26]. In our previous study, we analyzed monophonic sounds bereft of spatial movement. Thus, while Space Effort does not entirely correlate to the spatial domain, we introduce the stereo field and polyphonic sound in our application

in order to introduce an '*attitude*' toward space that the Laban Effort Factor suggests.

For example, chorus is a common effect used to thicken a sound through pitch and timing variations of a split/copied signal. Short delay times ($\leq$ 25 ms) keeps localization intact due to the precedence effect [27], and just like in audio production applications, additional panning of the effected sound will further spread the sound source image. Space Effort movements Direct and Indirect correspond to the attention a movement pays to the environment (which include spatial indicators), so that a chorus effect naturally underscores sonic spatial indications, including direct sound (doubled with no effect) and indirect sound (spatially diffuse). Space Effort builds a basic Chorus effect by controlling simple detuning, delay, and panning parameters. Indirect Space Effort detunes (half-semitone), delays (25 ms) and pans (50% right) a signal, whereas Direct Space Effect adds no effect.

The combination of synthesis parameters to Effort Factors is compiled as schematic in Table 3 and sound examples for the Eight Basic Effort Actions is shown in Table 4, which can be heard in [28].

5. DISCUSSION

Our sound application generates sounds based upon perceptual links between sound and movement. Individual sounds correspond to individual movements, which serve as basic building blocks of multi-modal stimuli (movement-sounds) and create the possibility for longer, multi-modal combinations of movement and sound. Effenberg [10] discusses the enhanced impact of multi-modal stimuli on motor perception, and Godøy [29] reasons that three motion-effort shapes: Impulsive, Sustained, and Iterative, informs *cross-modal* perception of sound and movement. The incorporation of peak emphasis into our design characterizes the shape of our sounds in accordance with two of Godøy's Shape descriptions: Impulsive and Sustained [29]. We did not find enough qualitative or quantitative evidence to help us construct the third sound shape, Iterative. We hypothesize that modulating signals are perceptible, and in an effort to further understand how Space Effort may be perceived in future study, we chose Space Effort to control simple frequency and timing modulations (chorus).

We recognize that our findings for peak amplitude emphasis does not include *when* in time the peak will occur. This problem may be due to that Time Effort Factor does not correlate to our general use of 'time' when describing sound. Specifically, Time Effort does not directly correspond to event timing or sonic duration. Instead, Time Effort describes the urgency of a movement, which can be Sudden or Sustained. Thus, our current solution is to manually align the 'peak' of a movement and the sound (a user parameter controlling peak emphasis location). We may be able to identify when a 'peak' occurs with respect to the movement trajectory by further studying the physical limits of robotic platforms.

We used qualitative results of Weight Effort to help bifurcate the differences between pitch and timbre. While this control stems from a qualitative finding, an exercise from a Laban/Bartenieff Movement Studies (LBMS) training workshop may help underscore the use of formant frequencies in relation to the body. In the LBMS exercise, participants are asked to vocalize vowel formants ($\bar{u}$ $\bar{o}$ $\bar{a}$ $\bar{e}$ $\bar{i}$), focusing on the resonance of the sound within different areas of the body. The vowels resonate from the lower abdomen [$\bar{u}$] up into the head cavity [$\bar{i}$]. The vertical ascent through the body moves due to a shift in Weight Effort, and may help explain qualitative results distinguishing pitch and timbre. The use of a formant filter in our synthesis application allows sounds to be shaped using frequency irrespective of pitch.

6. CONCLUSION

Our synthesized sounds parallel single movement trajectories, currently lasting between one to four seconds. Indeed, the duration of sounds we generate falls within the perceptual timing of Godøy's *meso-level* timescale, a timing range that suggests the amount of information that may be retained in short-term memory [30]. Generating longer and more complicated, multi-tier movements with controllable movement qualities is an area of ongoing research. Using our application, we could feasibly sequence movements and sounds together to form multi-movement phrases and patterns. The development of longer phrases would then allow for the study of Laban's fourth Effort Factor: Flow, which roughly speaking, describes movement phrasing. Further investigation into longer movement phrases may also be necessary to consider Space Effort quantitatively.

In order to develop a more robust sonificaiton framework, further research into the known physical limits of robotic platforms (i.e. gears and motors) may help deduce sonic duration as well as timing for peak amplitude emphasis. Our proposed framework provides a basis for mobile real-time movement sonification. An upcoming perceptual study using these designed sounds will further unpack related issues, and the study will help document how empirically informed sound design may be used to improve the perception of expressive robotic movement.

7. ACKNOWLEDGMENT

The study was made possible in part by the University of Virginia Data Science Institute, Office of the President, and the University of Virginia Office of Graduate and Postdoctoral Affairs. The authors also wish to thank Catherine Maguire, CMA, for movement training in the Laban Effort Factors.

8. REFERENCES

[1] A. LaViers, L. Bai, M. Bashiri, G. Heddy, and Y. Sheng, "Abstractions for design-by-humans of heterogeneous behaviors," in *Dance Notations and Robot Motion*. Springer Tracts in Advanced Robotics (STAR), 2015, pp. 237 – 262.

[2] H. Knight and R. Simmons, "Expressive motion with x, y and theta: Laban effort features for mobile robots," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014, pp. 267 – 273.

[3] L. Bai, J. Bellona, L. Dahl, and A. LaViers, "Design of perceptually meaningful quality in robotic motion," in *Workshop on Artistically Skilled Robots, IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016.

[4] L. Dahl, J. Bellona, L. Bai, and A. LaViers, "Data-driven design of sound for enhancing the perception of expressive robotic movement," in *International Conference on Movement and Computing*, 2017.

[5] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *International Conference on Digital Audio Effects*, 2007, pp. 237–244.

Table 3: Sound controls for Laban's Effort Actions leveraged from [4], as shown in table 2

Effort Factor	Parameter	Quality	Quality
		Strong	Light
Weight	Peak Envelope Emphasis	0 dBFS	-3dBFS
Weight	Overall Amplitude	0 dBFS	-6dBFS
Weight	Oscillator Frequency	C3 (130.81 Hz)	C5 (523.25 Hz)
Weight	Filter Formant Frequency	Eb4 (311.13 Hz)	G#7 (3322.4 Hz)
		Sudden	Sustained
Time	Peak Envelope Emphasis	-3dBFS Peak	-18dBFS No Peak
Time	Low-Pass Filter Cutoff	10000 Hz	1000 Hz
Time	Pitch Amplitude	-12dBFS	0dBFS
Time	Noise Amplitude	0dBFS	-12dBFS
		Direct	Indirect
Space	Chorus:Pitch	No Effect	Half-semitone
Space	Chorus:Delay	No Effect	25ms
Space	Chorus:Pan	No Effect	Right 50%

Table 4: Sound qualities for Eight Basic Effort Actions based upon quantitative sound design

Movement	Tone	Volume	Amp. Envelope
Gliding	Bright Tone	Quieter	Sustained
Pressing	Dark Tone	Louder	Sustained, Small Peak
Floating	Dark Noise	Quieter	Sustained
Wringing	Dark Noise	Louder	Sustained, Small Peak
Dabbing	Bright Tone	Quieter	Peak Envelope
Thrusting	Dark Tone	Louder	Large Peak Envelope
Flicking	Bright Noise	Quieter	Peak Envelope
Slashing	Bright Noise	Louder	Large Peak Envelope

[6] M. M. Wanderley, "Quantitative analysis of non-obvious performer gestures," in *Gesture and Sign Language in Human-computer Interaction*, I. Wachsmuth and T. Sowa, Eds. Berlin: Springer Verlag, 2002, pp. 241 – 253.

[7] B. Buck, J. Macritchie, and N. J. Bailey, "The interpretive shaping of embodied musical structure in piano performance." *Empirical Musicology Review*, vol. 8, no. 2, pp. 92 – 119, 2013.

[8] A. O. Effenberg, "Movement sonification: Effects on perception and action," *IEEE MultiMedia*, vol. 12, no. 2, pp. 53 – 59, 2005. [Online]. Available: http://ieeexplore.ieee.org/document/1423934/

[9] G. Schmitz, B. Mohammadi, A. Hammer, M. Heldmann, A. Samii, and T. F. Mnte, "Observation of sonified movements engages a basal ganglia frontocortical network." *BMC Neuroscience*, vol. 14, no. 1, pp. 1 – 11, 2013.

[10] A. O. Effenberg, U. Fehse, G. Schmitz, B. Krueger, and H. Mechling, "Movement sonification: Effects on motor learning beyond rhythmic adjustments." *Frontiers in Neuroscience*, pp. 1 – 18, 2016.

[11] P. M. Vinken, D. Krger, U. Fehse, G. Schmitz, H. Brock, and A. O. Effenberg, "Auditory coding of human movement kinematics." *Multisensory Research*, vol. 26, no. 6, pp. 533 – 552, 2013.

[12] G. Schmitz and A. O. Effenberg, "Perceptual effects of auditory information about own and other movements," in *Proceedings of the 18th International Conference on Auditory Display*, Atlanta, GA, June 2012.

[13] H. Brock, G. Schmitz, J. Baumann, and A. O. Effenberg", "If motion sounds: Movement sonification based on inertial sensor data," *Procedia Engineering*, vol. 34, pp. 556 – 561, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877705812017080

[14] A. LaViers and M. Egerstedt, "Style-based robotic motion," in *Proceedings of the 2012 American Control Conference*, Montreal, QC, June 2012.

[15] J. Françoise, S. Fdili Alaoui, T. Schiphorst, and F. Bevilacqua, "Vocalizing dance movement for interactive sonification of laban effort factors," in *Proceedings of the 2014 conference on Designing interactive systems*. ACM Press, 2014, pp. 1079–1082. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2598510.2598582

[16] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities." *PLoS ONE*, vol. 8, no. 12, pp. 1 – 28, 2013.

[17] J. Wilson-Bokowiec and M. A. Bokowiec, "Kinaesonics: The intertwining relationship of body and sound." *Contemporary Music Review*, vol. 25, no. 1/2, pp. 47 – 57, 2006.

[18] J. Bellona, "Casting (for kinect, kyma, and processing)," http://jpbellona.com/work/casting/, 2013.

[19] K. Studd and L. Cox, *Everybody is a Body*. Indianapolis, IN: Dog Ear Publishing, 2013.

[20] R. Laban and F. C. Lawrence, *Effort: Economy of Human Movement*, 2nd ed. Macdonald and Evans, June 1974.

[21] J. Bellona, "Jpb.mod*, a data modification toolkit," http://jpbellona.com/work/jpb-mod/.

[22] I. Titze, T. Riede, and T. Mau, "Predicting achievable fundamental frequency ranges in vocalization across species." *PLoS Computational Biology*, vol. 12, no. 6, pp. 1 – 13, 2016.

[23] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *Journal of the Acoustical Society of America*, vol. 5, pp. 82 – 108, 1933.

[24] J. Catford, *A Practical Introduction to Phonetics (Oxford Textbooks in Linguistics)*, 2nd ed. Oxford University Press, 2002.

[25] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *The Journal of the Acoustical Society of America*, vol. 24, no. 2, 1952. [Online]. Available: http://asa.scitation.org/doi/pdf/10.1121/1.1906875

[26] M. L. Johnson and S. Larson, "Something in the way she moves – metaphors of musical motion," *Metaphor and symbol*, vol. 18, no. 2, pp. 63–84, 2003.

[27] B. Krapalos, M. R. M. Jenkin, and E. Milios, "Auditory perception and spatial (3d) auditory systems," *Technical Report CS-2003-07*, 2003.

[28] J. Bellona, L. Dahl, L. Bai, and A. LaViers, "Empirically informed sound synthesis application playing laban's eight basic effort actions," YouTube, 2017. [Online]. Available: https://youtu.be/R8NQ8LYfv08

[29] R. I. Godøy, M. Song, K. Nymoen, M. R. Haugen, and A. R. Jensenius, "Exploring sound-motion similarity in musical experience," *Journal of New Music Research*, vol. 0, no. 0, pp. 1–13, 2016. [Online]. Available: http://dx.doi.org/10.1080/09298215.2016.1184689

[30] R. I. Godøy, "Gestural affordances of musical sound," in *Musical gestures: sound, movement, and meaning*, R. I. Godøy and M. Leman, Eds. New York: Routledge, 2010, ch. 5, pp. 103–125.

INFLUENCES OF VISUAL AND AUDITORY DISPLAYS ON AIMED MOVEMENTS USING AIR GESTURE CONTROLS

Jason Sterkenburg, Steven Landry, Myounghoon Jeon

Mind Music Machine Lab
Michigan Technological University,
1400 Townsend Ave., Houghton, MI USA
{jtsterke,sglandry,mjeon}@mtu.edu

ABSTRACT

With the proliferation of technologies operated via in-air hand movements, e.g. virtual/augmented reality, in-vehicle infotainment systems, and large public information displays, there remains an open question about if/how auditory displays can be used effectively to facilitate eyes-free aimed movements. We conducted a within-subjects study, similar to a Fitts paradigm study, in which 24 participants completed simple aimed movements to acquire targets of varying sizes and distances. Participants completed these aimed movements for six conditions – each presenting a unique combination of visual and auditory displays. Results showed participants were generally faster to make selections when using visual displays compared to displays without visuals. However, selection accuracy was similar for auditory-only displays when compared to displays with visual components. These results highlight the potential for auditory displays to aid aimed movements using air gestures in conditions where visual displays are impractical, impossible, or unhelpful.

1. INTRODUCTION

Air gesture control – the operation of devices by in-air hand movements – has potential to empower users with a natural and rich level of control over their devices. Auditory displays, in combination with gesture controls, could improve technology accessibility for visually-impaired users and allow for eyes-free interaction for sighted users. However, it is unknown how auditory displays affect aimed movements using air gesture controls. Auditory display design is particularly interesting in application to air gesture controls because, unlike many other forms of technology, air gesture controls allow for continuous tracking of user hand positions. Currently, little is known about how different sonification strategies may affect aimed movement performance when using air gesture controls.

Some studies have investigated aimed movement performance using air gesture controls [e.g., 8-9] and even explored the concept of eyes-free aimed movements [9] using only kinesthetic information. Other studies have examined the impact of auditory displays on target acquisition performance [1-2]. However, to our knowledge there is little to no existing literature exploring the utility of auditory displays in conjunction with air gesture controls in aiding target acquisition tasks. Most existing literature surrounding the topic of auditory displays and air gestures have focused on target localization, i.e., finding the point of origin of a sound in space [e.g., 3-7].

We conducted an experiment to learn how auditory displays affect aimed movement performance using air gesture controls. We made comparisons between two sonification strategies: (1) a discrete auditory display – playing a sound whenever the user is on the target and (2) a continuous auditory display – playing sound continuously from the start of the movement until selection, and playing a discrete sound when the user is on target. We also made comparisons among auditory-only, visual-only, and visual-auditory displays, as well as a control condition for which there was no visual or auditory display.

2. METHODS

2.1. Design Guidelines

Soukoreff and MacKenzie [3] wrote a paper outlining several guidelines which supported the ISO 9241-9 standards for the evaluation of pointing devices in human-computer interaction. In keeping with standard evaluation of pointing devices, we followed each of those standards as much as possible. This is our justification for (1) our use of the Shannon formulation of index of difficulty, (2) our range of movement difficulties, (3) our adjustments for selection accuracy, (4) and our calculation of throughput.

2.2. Apparatus

We used a LEAP Motion as our hand-position tracking sensor and we used Pure Data – an open source graphical programming language – to develop our target selection task (Figure 1). As the participant moves their hand above the sensor, a cursor matches the position of the person's hand along the x-axis (no y-axis data were recorded) and makes corresponding movements on the screen. All cursor movements were mapped one-to-one to hand movements.

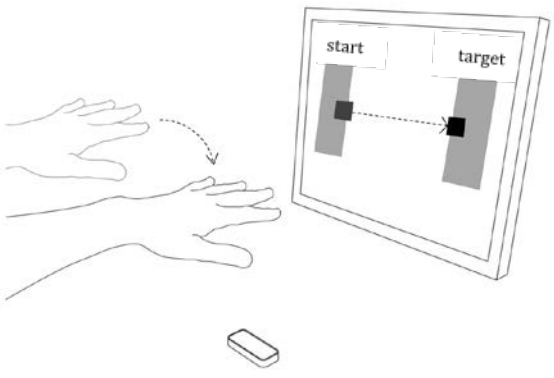


Figure 1: Illustration of experimental setup.

https://doi.org/10.21785/icad2017.065

2.3. Participants

A total of 24 undergraduate psychology students were recruited to complete our study (Table 1). All participants were given course credit as compensation for their participation. Only one person reported having experience using a LEAP Motion before.

Age (yrs)	Gender	Handedness
Mean=19.75	Males:14	Right:21
SD=1.96	Females:10	Left:3

Table 1: Demographic statistics for participants

2.4. Experimental design

We used full factorial within-subjects design with a total of six conditions:

AC – continuous auditory display
AD – discrete auditory display
VAC – visual plus continuous auditory
VAD – visual plus discrete auditory
V – only visual display
Control – visual removed upon start, no audio

2.5. Sound design

There were two different sound designs: a discrete auditory display and a continuous auditory display. The discrete auditory display consisted only of a pink noise that played as long as the cursor is within the target. The continuous auditory display constantly plays a sine wave that increases in frequency as the cursor gets closer to the target. The pitch increases as a function of the square (x^2) of current fraction of the total distance to the target that the cursor has traveled (Equation 1). The pitch increases one octave from the start to the target position. The continuous auditory display also played a pink noise when the cursor was within the target position.

$$Pitch_t = 440 + \left(\sqrt{440} * \left(1 - \left(\frac{|position_t - target\ position|}{target\ position}\right)\right)\right)^2 \quad (1)$$

2.6. Procedure

2.6.1. Practice

After providing informed consent and filling out a brief demographic survey, participants were first introduced to the general purpose of the experiment and given five minutes of guided practice during which they were exposed to each of the six different conditions. Participants were seated in a chair in front of a computer and a leap motion fixed at a 45 degree angle to the table. Participants were able to complete the task with their left or right hand but they were asked to not switch hands during the experiment. Participants were encouraged to take breaks between selections or conditions as needed.

2.6.2. Testing

After selecting the start button (open hand = select gesture), a target appeared somewhere to right on the screen. For visual conditions, participants can see the cursor and target, which changes color when the cursor enters it. For non-visual conditions the cursor and target are not visible. For each of the six conditions participants completed a total of 48 selections, 12 for each of 4 difficulty levels (ID = 2, 3, 4, 5). Each condition took about 6-8 minutes – the experiment lasted about an hour overall.

2.7. Statistics

Repeated-measures ANOVAs were conducted to identify differences between conditions. Two-tailed, paired-samples t-tests were conducted. A Holm-Bonferroni correction was used to decrease the number of Type-1 errors.

3. RESULTS

3.1. Selection time

Repeated measures ANOVA results indicate main effects for condition, $F(5,19) = 36.4$, $p < .001$, as well as difficulty $F(3,21) = 14.9$, $p < .001$. There was also a significant interaction, $F(15,545) = 3.83$, $p < .001$,which can be seen as a difference in slope of the lines in Figure 2. Paired comparisons (Table 2) showed participants were slower to make selections when using continuous (AC) and discrete (AD) auditory displays compared to conditions with visual displays and the control condition (Figure 2).

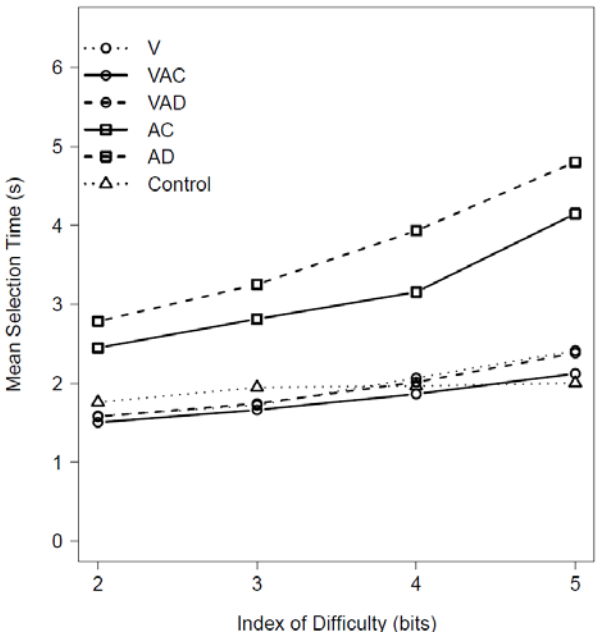


Figure 2: Average selection times for each condition across difficulty levels.

	AC	AD	Control	V	VAC
AD	<.001*	--	--	--	--
Control	<.001*	<.001*	--	--	--
V	<.001*	<.001*	1.00	--	--
VAC	<.001*	<.001*	.011	.057	--
VAD	<.001*	<.001*	1.00	1.00	.107

Table 2: P-values for pairwise comparisons of average selection times.

3.2. Selection accuracy

3.2.1. Error

Repeated measures ANOVA results show a main effect by condition, $F(5,19) = 34.4$, $p < .001$. Difficulty also showed main effects, $F(3,21) = 54.3$, $p < .001$, as well as an interaction with condition, $F(15,345) = 10.0$, $p < .001$, which can be seen by the difference in slopes of lines in Figure 3. Paired comparisons (Table 3) showed that participants' selection error, defined by the absolute value of the distance between the final cursor position and the closest edge of the target, was significantly higher for the control condition (Figure 3) compared to all other conditions. These tests also revealed that the discrete auditory display (AD) led to significantly higher error compared to all conditions other than the control. The AC condition led to significantly higher error compared to all conditions other than AD, VAD, and Control. All other conditions were statistically equivalent.

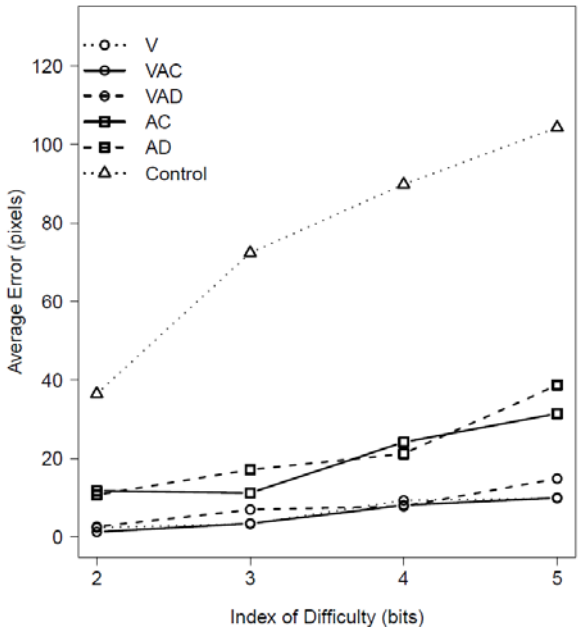


Figure 3: Average adjusted error for each condition across difficulty levels.

	AC	AD	Control	V	VAC
AD	1.00	--	--	--	--
Control	<.001*	<.001*	--	--	--
V	0.007*	0.002*	<.001*	--	--
VAC	0.006*	0.001*	<.001*	1.00	--
VAD	0.019	0.005*	<.001*	1.00	1.00

Table 3: P-value for pairwise two-tailed t-tests for selection error across conditions. * indicates a statistically significant difference.

3.2.2. Percent correct

ANOVA results showed a main effect for condition, $F(5,19) = 67.1$, $p < .001$. Figure 4 shows that conditions appear to largely be similar with the exception of the control condition which is significantly lower, which can be seen in Table 4. Difficulty also showed a main effect, $F(3,21) = 472$, $p < .001$, which is especially obvious in Figure 4. There was also an interaction between condition and difficulty, $F(15,345) = 2.99$, $p < .001$. There appears to be some separation between visual and non-visual displays at higher difficulties. Possible explanations will come in the discussion section.

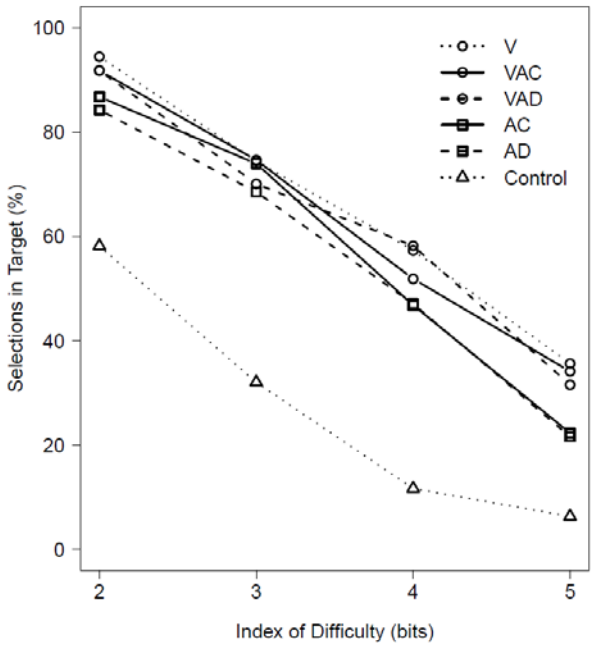


Figure 4: Average percent correct across conditions for each difficulty.

	AC	AD	Control	V	VAC
AD	1.00	--	--	--	--
Control	<.001*	<.001*	--	--	--
V	0.57	0.25	<.001*	--	--
VAC	0.76	0.43	<.001*	1.00	--
VAD	0.89	0.57	<.001*	1.00	1.00

Table 4: P-values for pairwise two-tailed t-tests for selection accuracy across conditions. * indicates a statistically significant difference.

3.3. Throughput

Throughput is a calculation that accounts for both the accuracy of the movement – difference between endpoint position and the center of the target – and movement time. This provides a measure of overall movement performance by information conveyed in bits per second. Repeated measures

ANOVA results showed significant differences between conditions $F(5,19) = 91.4$, $p < .001$ for throughput. Paired comparisons showed that participants had higher throughput with visual displays (VAC, VAD, V) compared to auditory-only conditions (AC, AD, Control) (Table 5). There was also a main effect for index of difficulty (ID), $F(3,21) = 35.0$, $p < .001$, and a statistical interaction, $F(15,345) = 2.04$, $p = .0122$.

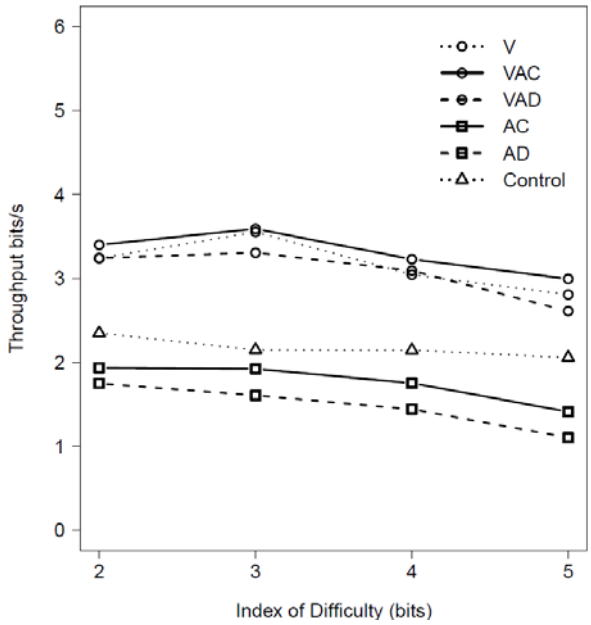


Figure 5: Average throughput across conditions for each difficulty.

	AC	AD	Control	V	VAC
AD	0.019	--	--	--	--
Control	<.001*	<.001*	--	--	--
V	<.001*	<.001*	<.001*	--	--
VAC	<.001*	<.001*	<.001*	.304	--
VAD	<.001*	<.001*	<.001*	.323	.047

Table 5: P-value results for pairwise two-tailed t-tests for throughput across conditions. * indicates a statistically significant difference.

4. DISCUSSION

These results convey a nuanced story about the influence of auditory displays on aimed movements using air gesture controls. As expected, visual displays resulted in faster movement times compared to auditory displays. Previous literature has shown that visual information is more readily integrated into trajectory corrections [10], suggesting that using auditory displays to convey information about movement trajectory is more effortful than with visual displays. One possible explanation for the difference in selection times between visual and auditory-only displays is that people are better able to accurately estimate the distance to the target and close the gap more quickly in the initial ballistic phase of movement when using visual displays, as opposed to the auditory-only displays which require more searching behavior. Regarding selection accuracy, however, auditory-only displays led to similar percentages of in-target selections compared to conditions with visual displays, especially at lower levels of difficulty. Interestingly, auditory-only displays consistently resulted in a statistical interaction, showing slower and less accurate movements, especially for much higher difficulty movements (ID = 4, 5). We suppose that the relatively poor performance for auditory-only displays for selection times may be because participants are receiving less information about the relative position of the cursor and the target, and as a result, need to make more fine motor corrections once they are close to the target.

Comparing between the continuous and discrete auditory-only displays, the continuous auditory display led to faster selection times and comparable accuracy, leading to overall higher throughput. The same pattern was not as clear when comparing continuous and discrete audio paired with visual displays (VAC and VAD), possibly as a result of participants deferring to visual information when it is available.

Overall, results indicate that auditory-only displays are not as effective as visual displays at guiding aimed movements in target acquisition tasks among sighted users. However, the data suggest that targets can be selected with similar levels of accuracy when using auditory-only displays, especially when movements are less difficult (ID = 2, 3). This suggests the potential for using auditory displays (continuous or discrete) for facilitating eyes-free target acquisitions using air gesture controls. For example, in vehicle contexts, auditory-only displays can result in the same accurate performance in the secondary gesture task, while maintaining visual attention on the road. Therefore, further applied research is required to identify the relationship among the task demand (e.g., level of difficulty), multi-modalities, and different types of auditory displays.

5. REFERENCES

[1] Hatfield, Brent C., William R. Wyatt, and John B. Shea. "Effects of auditory feedback on movement time in a Fitts task." *Journal of motor behavior* 42.5 (2010): 289-293.

[2] de Grosbois, John, Matthew Heath, and Luc Tremblay. "Augmented feedback influences upper limb reaching movement times but does not explain violations of Fitts' Law." *Frontiers in psychology* 6 (2015).

[3] Soukoreff, R. William, and I. Scott MacKenzie. "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI." *International journal of human-computer studies* 61.6 (2004): 751-789.

[4] Pierno, Andrea C., Andrea Caria, and Umberto Castiello. "Comparing effects of 2-D and 3-D visual cues during aurally aided target acquisition." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46.4 (2004): 728-737.

[5] Pierno, Andrea C., et al. "Effects of increasing visual load on aurally and visually guided target acquisition in a virtual environment." *Applied ergonomics* 36.3 (2005): 335-343.

[6] Zahariev, Mihaela A., and Christine L. Mackenzie. "Auditory contact cues improve performance when

grasping augmented and virtual objects with a tool." *Experimental brain research* 186.4 (2008): 619-627.

[7] Marentakis, Georgios, and Stephen A. Brewster. "A study on gestural interaction with a 3d audio display." *International Conference on Mobile Human-Computer Interaction*. Springer Berlin Heidelberg, 2004.

[8] Grossman, Tovi, and Ravin Balakrishnan. "Pointing at trivariate targets in 3D environments." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2004.

[9] Cockburn, Andy, et al. "Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback." *International Journal of Human-Computer Studies* 69.6 (2011): 401-414.

[10] Elliott, Digby, Werner F. Helsen, and Romeo Chua. "A century later: Woodworth's (1899) two-component model of goal-directed aiming." *Psychological bulletin* 127.3 (2001): 342.

ROTATOR: FLEXIBLE DISTRIBUTION OF DATA ACROSS SENSORY CHANNELS

Juliana Cherston

Responsive Environments Group
MIT Media Lab
75 Amherst Street
Cambridge MA, 02139 USA
cherston@media.mit.edu

Joseph A. Paradiso

Responsive Environments Group
MIT Media Lab
75 Amherst Street
Cambridge MA, 02139 USA
joep@media.mit.edu

ABSTRACT

'Rotator' is a web-based multisensory analysis interface that enables users to shift streams of multichannel scientific data between their auditory and visual sensory channels in order to better discern structure and anomaly in the data. This paper provides a technical overview of the Rotator tool as well as a discussion of the motivations for integrating flexible data display into future analysis and monitoring frameworks. An audio-visual presentation mode in which only a single stream is visualized at any given moment is identified as a particularly promising alternative to a purely visual information display mode. Auditory and visual excerpts from the interface are available at **http://resenv.media.mit.edu/rotator**.

1. INTRODUCTION

While walking down the streets of a bustling city, we are unfailingly submerged in a three-dimensional, 360° sonic cacophony of urban life that is commonly referenced in academic literature as an *auditory scene*. Our eyes, cued by the sounds around us, dart purposefully around the scene. For example, when a car screeches, we impulsively jolt our heads to bring the source of the sudden noise into our field-of-view. In this way, we are accustomed to using visual and sonic information in tandem, placing our eyes where they are most beneficial and tuning our attention to precise features embedded within an auditory scene. However, unlike audio, visual input is directional, providing a field-of-view of only 130-135° vertically and 200° horizontally with respect to the center of one's gaze [1]. It is therefore worth considering whether the brain's natural methods for distributing data across our auditory and visual sensory channels can aid in the design of a productive high-dimensional data analysis framework.

2. OVERVIEW OF ROTATOR TOOL DESIGN

The Rotator tool is a web-based data analysis framework that enables flexible distribution of scientific data between the user's eyes and ears. It is designed to study how our perception of the structure of a dataset can be informed and potentially optimized by deliberate manipulation of the presentation modes of its constituent data streams. The expected workflow is as follows: a user loads in a dataset to be analyzed. The streams within the dataset are assumed to have a geometric interpretation and are displayed on a suitable map. For example, the data may be derived from a physical sensor network or may represent different stages of a process that can be laid out as a schematic. The user begins to explore the data by moving and resizing auditory and visual windows around the node network (see Figure 1 as an example). Nodes falling within one of these windows will be displayed accordingly, either as spatialized audio (where sonified streams are distributed with respect to the center of the sliding audio window), or as a series of line/scatter plots rendered adjacent to the node map. The user may find that listening to a spatialized sonification of all of the data nodes while only looking at visual representations of a few nodes at a time is most comfortable (Figure 1, center). Alternatively, the user may wish to divide the available data streams between his or her two sensory modes (Figure 1, right). When both auditory and visual modes are enabled, a vertical red line slides across each visual plot to provide an approximate indication to users of the data region currently being sonified (Figure 5, left column).

In order to further explore the data, the user can adjust the data-to-audio mapping settings in the audio control panel (Figure 2), shift the range of data within each stream that is currently visualized or sonified, view the FFT of any individual data stream, enable a 'play' mode that simulates a real-time monitoring scenario, or cluster the data nodes, among other available controls.

3. THEORY AND PRIOR ART

3.1. Multichannel Monitoring and Expectancy Violation

Rotator relies on the human brain's capacity for auditory stream segregation and multi-channel visual monitoring. However, studies have repeatedly shown that humans can only monitor four to five items simultaneously before experiencing cognitive overload [2, 3]. Generally, research addressing this perceptual limitation centers on methods for pre-processing, filtering, and auto-labeling data in order to render only the most critical information to the user. That said, many such commonplace approaches, e.g. dimensionality reduction algorithms, inherently obscure the meaning of each individual data stream. It is therefore useful to recognize that the visual system and the auditory system can serve complimentary roles in data display. For example, while visual displays work well for providing detailed renderings of local areas of a dataset, aural scanning is better suited for finding regions of interest with short temporal durations in a large dataset. In each case, the user is particularly sensitive to anomaly. For example, in audio, Cariani

https://doi.org/10.21785/icad2017.009

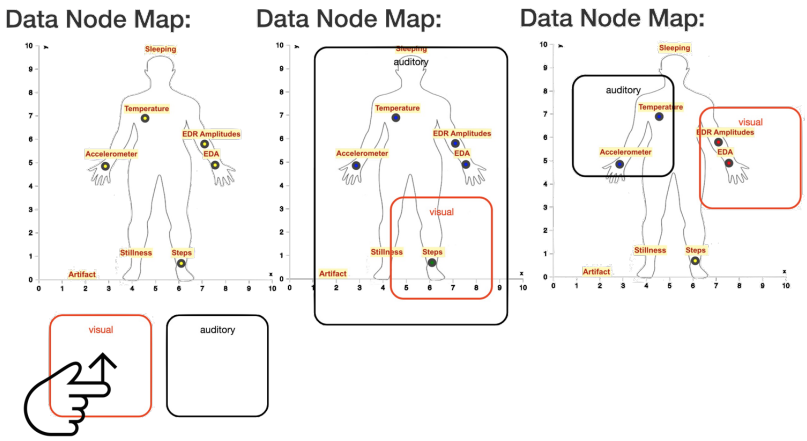


Figure 1: Three sample configurations for Rotator's 'Data Node Map' interface applied in this case to data from a wearable biosensing system. Users resize and slide auditory and visual windows around a map of data stream nodes in order to dictate in which sensory channel each node will be displayed

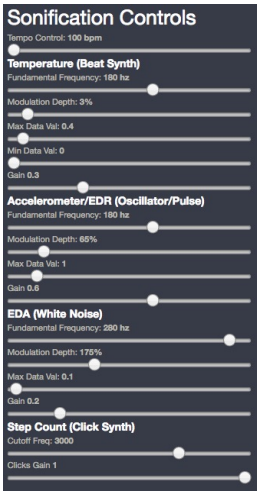


Figure 2: Sample configuration for Rotator's left-hand sonification control panel pictured for the wearable biosensing system application area. The user can adjust the overall tempo as well as data-to-audio mapping parameters for each synthesizer. For example, in the case of the beat effect synthesizer used to sonify body temperature, the user can modify fundamental frequency, modulation depth, min/max data values expected in the stream, and gain

and Micheyl write that the human benefits from expectancy violation effects in the auditory cortex: 'On all timescales, repeating patterns of sound and their evoked auditory events build up strong representational expectancies of their continuation. This effect is created even with arbitrary and highly artificial repeating sound patterns' [4]. In Rotator, users can probe expectancy-violating auditory effects in the visual domain and vice-versa. Expectancy violating auditory events may include the fleeting increase in frequency of a beat effect synthesizer, or the pulsation of an envelope synthesizer, to name some examples.

3.2. Auditory Scene Analysis and Stream Segregation

Auditory Scene Analysis (ASA) describes a set of heuristics that model the organization of incoming auditory data into a scene around the listener. Within an auditory scene of discrete and sufficiently segregated sound sources, the listener's attention can shift from source to source [5]. Rotator draws from prior work in stream segregation, notably in the heavy use of spatialization, pitch segregation, and the overlay of continuous and intermittent sounds.

Pioneering work on auditory scene analysis was conducted by Bregman in 1994 [6]. Cariani's biological framing of 'expectancy violation' in the previous section is qualitatively similar to Bregman's 'old-plus-new' heuristic: when new components are abruptly added to a spectrum of partials, the ASA system is skilled at deducing which partials are a continuation of the previous signal and which are newly added. The newly added partials are perceived as a separate sound [6].

Efforts have been made to model perceptual segregation of streams in terms of common auditory parameters. For example, the van Noorden Diagram maps the perception of two tones with respect to their pitch difference and intertone onset interval into regions where one or two streams are perceived, as well as an ambiguous region in which perception is dependent on attention (diagram available in [7]). While the van Noorden diagram is typically used to map stream segregation with respect to pitch differences, Barass et al. have broadened the palette of such diagrams to other audio properties including brightness of a noise grain, and then to amplitude and inter-level difference panning of the noise grain [5]. van Noorden diagrams can be useful in guiding multi-stream sonification algorithms away from regions of perceptual ambiguity.

Spatialization can be used to minimize interference and masking effects in audio. The use of spatialization to regulate stream segregation has a number of specific advantages. Firstly, it is an independent parameter in the sense that the spatialization of a particular stream can be modified without risk of the change interfering with other parameters (conversely, if e.g. pitch and timbre are modified independently, interaction between these two dimensions may cause ambiguity in the resulting stream) [8]. Secondly, spatialization provides the user with a physical map for the sound sources which can aid in the sonic learning process.

The capacity for spatialized audio to aid in the perception of segregated streams has been evaluated only a handful of times in prior work. A study that is most closely related to the current discussion compared user identification of two pitch-segregated signals that are spatially segregated with two pitch-segregated signals emanating from a single source. Researchers found improved signal classification accuracy when audio streams are spatially segregated, among other findings described in [9].

3.3. Data Sonification Platforms

Up until now our discussion has focused predominantly on theory and prior work in the study of auditory perception. Rotator, in more practical terms, is intended as a sonification platform.

A number of generalized data sonification platforms have been developed in the past. Four examples are the Sonification Sandbox (a Java application developed by the Georgia Institute of Technology) [10], SonArt (a platform developed at Stanford University) [11], Personify (a scientific data sonification platform built at CSIRO) [12], and MUSE (A musically-driven data sonification platform created by UC Santa Cruz) [13].

Each tool makes use of visual information to aid in the user experience. For example, Personify asks users to customize a visual representation of a data-to-audio mapping space where axes correspond to musical properties, and the Sonification Sandbox allows users to view line plots of data streams as they are being sonified. However, unlike the sonification platforms cited above, the Rotator platform is specifically aimed at diversifying the way that users *distribute* data across their senses, which we have yet to see as the focus area of a sonification tool.

4. SYSTEM ARCHITECTURE

4.1. Architectural Overview

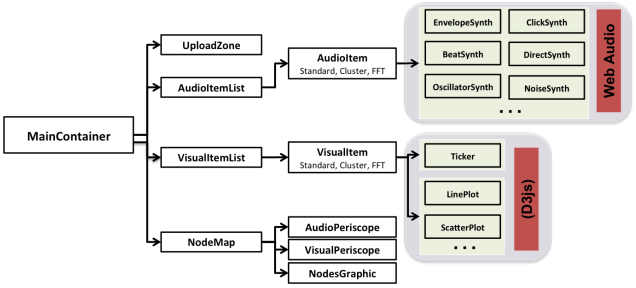


Figure 3: Simplified React view component hierarchy for the Rotator application. Note that additional communication loops exist that are not pictured

Rotator is a client-side application built in Javascript and architected in React and Flux. The most heavily used low-level Javascript libraries are D3 (for visualization) and Web Audio (for sonification). The decision to build Rotator as a web-based application was reached for a few principle reasons. Firstly, real-time, web-based audio technology has improved tremendously over the last decade and is rapidly becoming a standard. Secondly, previous experience integrating together common audio processing tools like Pure Data, Max/MSP, and Ableton Live suggest that the installation process for a new user can be cumbersome; a web-based application requires only a browser. Finally, extending a niche visualization tool such as ROOT[1] was entertained as a possibility but would target a narrower user base than a web-based tool. Web Audio does have some drawbacks, however. Most notably, it is still under active development and therefore lacks certain basic audio streaming features that are commonplace in desktop audio processing software, e.g. adjustable sampling rate. Secondly, Web Audio is yet to be fully ported to React and must therefore be integrated by the developer. (Note that [15] is a first pass at Web Audio integration into React, but is yet to support audio spatialization).

Figure 3 shows a simplified flowchart of the React view components designed for the Rotator application. There are three principle data flow chains: one controlling sound synthesis, one controlling visualizations, and one controlling the node map used to navigate within a dataset. A combination of callback functions and Flux data store updates enables communication between sibling components. Further, a custom audio scheduler was built for audio synchronization across Rotator components.

4.2. Synthesizer Architecture

In the current software architecture, each data stream's associated synthesizer component is mounted and activated when the user wishes to sonify the stream. Bearing in mind principles in psychoacoustic theory described in Section 3, six synthesizer types have been built as default options: **noise**, **envelope**, **clicks**, **oscillator**, and **beats**, and **direct**. Each synth's parameters are controlled via user-defined values supplied in the lefthand control panel previously pictured in Figure 2. In addition, it is possible for a user to integrate an entirely new audio synthesizer component as long as it accepts a set of React props required by all Rotator synthesizers.

The white noise synth populates an audio buffer of size 100,000 with random values and then channels the buffer through a low-pass filter with a cutoff frequency driven by the data. The envelope synthesizer defines an envelope's attack, decay, sustain, and release times, ramping up an oscillator to a frequency driven by the incoming data. The click synth is another iteration of the envelope synth but with parameters shorter in duration, and applied to a buffer filled with white noise. The oscillator synth uses incoming data to modulate the pitch of the oscillator. The beat synth creates a beat effect between two oscillators, using data to modulate the small frequency gap between each oscillator. The further that the incoming data is from a preset threshold value, the higher the frequency of the beats. Basic wave shaping distinguishes data values above and below the threshold; values below the threshold trigger two sinusoidal oscillators, and values above the threshold trigger two triangle oscillators. Finally the direct synth creates a direct audification of the data by populating an audio buffer that is played back at a rate determined by the user-defined tempo. Audio samples can be accessed at **http://resenv.media.mit.edu/rotator**.

5. APPLICATION AREAS

Three application areas for Rotator were implemented. Two scenarios (quantum algorithm interpretation, temperature monitoring in an experiment) will be briefly described. The most promising scenario (biosensor data analysis) is described more thoroughly and is incorporated into a user study described in Section 6. Figure 4 shows the node maps used while testing each scenario.

[1]ROOT is a C++ visualization framework used in particle physics [14].

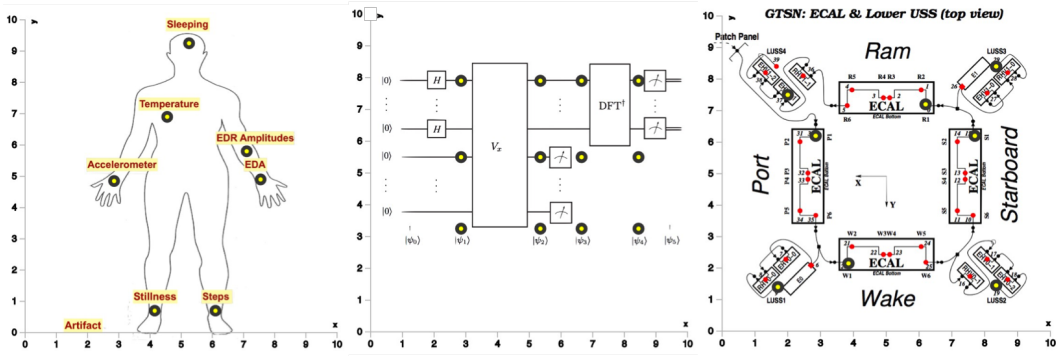


Figure 4: Auditory and visual windows can be slid and resized across appropriate node maps, as in Figure 1. Maps for the three application areas considered are pictured: biosensor data (left) quantum algorithm circuit (center) temperature monitoring in an experiment (right)

5.1. Quantum Algorithm Interpretation

Audio may be a very suitable data display mode for quantum systems. This presumption is due in large part to a phenomenon known as quantum superpositon, which describes the capacity for a quantum bit to exist in two states at once (with some probability of a measurement yielding either state). Similarly, as described in section 3.2, audio can be layered in a way that triggers the simultaneous perception of multiple streams, and is thus a natural display mode for data encoded in a quantum system. Rotator was used to sonify the quantum states that arise in a famous quantum factoring algorithm called Shor's algorithm [16]. We were specifically intrigued by the possibility for Rotator to immerse a user in different stages of the algorithm simultaneously. For instance, using Rotator, it is possible to hear an algorithm state prior to a quantum operation in one's left ear and an algorithm state following quantum operation in one's right ear, while visually examining plots of the remaining stages of the algorithm. Furthermore, a quantum algorithm's circuit diagram provides a natural node map within Rotator that can guide user interaction.

Shor's algorithm is an interesting choice for sonification. On the one hand, it is a quintessential quantum algorithm with a periodic structure that can be elucidated sonically. Oscillator synths were used to convey periodic states and click synths were used to convey the results of quantum measurements on the states. On the other hand, the high quantum bit numbers per register that are required even for factoring small N result in a very large superposition of states to work with. The states have equal probability amplitude at many stages of the algorithm (as compared to e.g. the quantum harmonic oscillator where a time-dependent term modulates the probability amplitudes as the system evolves), and the important periodic structure in the algorithm does not lie in the probability amplitudes of the states, but instead in the series of states themselves. To make this periodicity apparent in audio, we must iterate through the states in superposition rather than play them simultaneously, which breaks from our initial motivation to sonify quantum systems in the first place! Be that as it may, the current sonification approach is more generalizable across classical algorithms. Furthermore, for very low N, it may be possible to sonify the states of all bits in a register simultaneously rather than iterate through each bit state, which is worthy of further consideration. However, to attain the full benefit of sonic display of quantum data, it is recommended that future approaches select quantum systems with small numbers of states and with a time dependent parameter that modulates the probability amplitudes of these states.

5.2. AMS Temperature Data

The alpha-magnetic spectrometer (AMS) is a physics experiment onboard the International Space Station (ISS) responsible for detecting particles that may contribute to our understanding of dark matter, anti-matter, and cosmic rays [17]. Currently, hundreds of sensor node readings are presented as tables in the AMS control room and are color coded in red, green, and yellow on the basis of expected vs. anomalous behavior.

A small, offline dataset consisting of 8 temperature nodes located throughout the detector was provided for consideration of alternative data display modes, and an initial sonification scheme was developed. Rather than aim for maximal stream segregation, an array of temperature nodes lends itself to representation as a soundscape that blends together, creating a feeling in the listener of being immersed in a heat map. Therefore, a simple mapping was created using eight spatialized oscillator synths, with temperature controlling the pitch of each synth. As users grow accustomed to this ambient hum, they will be able to identify warmer and cooler regions of the detector while growing sensitive overtime to perceived anomalous behavior, which can then be validated visually. Rotator's node clustering feature was of particular use in this application area, since each region of the detector behaves similarly.

The AMS temperature dataset is particularly appealing due to its behavior at multiple timescales: high-frequency oscillations are caused by 92-minute orbitals, and low frequency oscillations are caused by gradual drift in ISS orbit. In order to perceive both scales at once using visual tools, it is necessary to either zoom in and out, else look back and forth between two plots at different time scales. A tool like Rotator can be used to provide the user a sonification of the high frequency oscillations of a temperature stream, while an accompanying visualization simultaneously provides the user with larger timescale, low frequency behavior. This approach is most readily accomplished in Rotator by increasing the modulation depth of a stream's synthesizer in order to amplify smaller-scale changes in the sound. Meanwhile, a lowpass-filtered signal can be fed in by the user for visualization. In this way, the user is able to simultaneously track variations of a temperature node at two different timescales, while only monitoring a single auditory and single visual track.

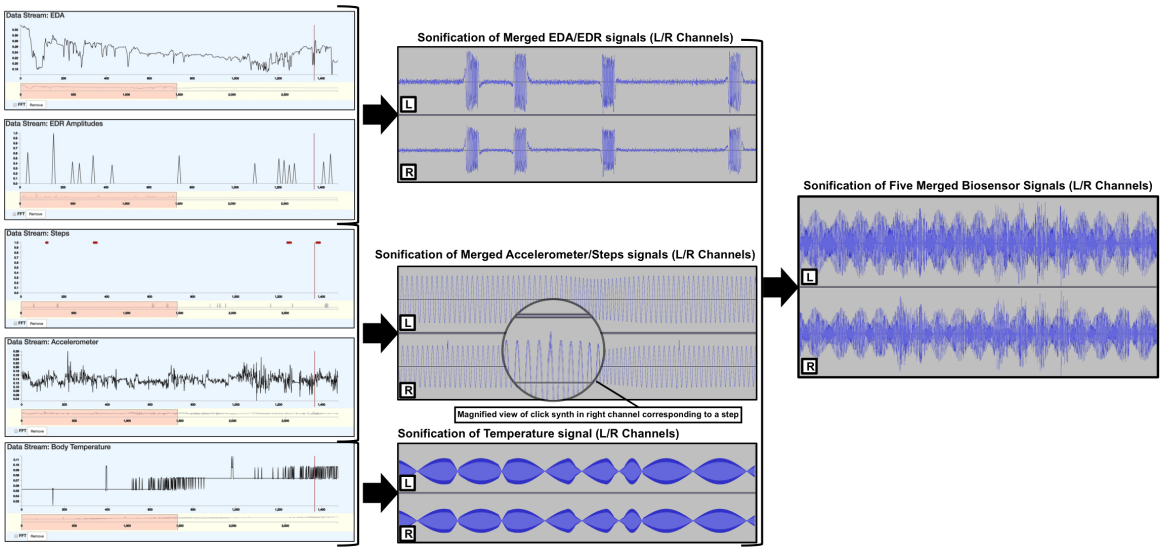


Figure 5: Sample sonification of five biosensor data parameters. The left-most five plots contain excerpts from each data stream as visualized in Rotator (note that the pink bar at the base of each plot is resizable and slidable to adjust data range displayed). Three audio streams depict audio from EDA and extracted EDRs (top), accelerometer and extracted steps (middle) and temperature (bottom). The rightmost plot shows the resulting audio stream when all five parameters are sonified at once.

5.3. Biosensor System Data

Data from one participant of a largescale user study on wellbeing was provided as a testbed for the Rotator tool (see [18] for study information). The participant of this study was equipped with the Affectiva Q wrist-worn physiological sensor that records acceleration, skin temperature, and Electrodermal Activity (EDA).

The principle task of the researchers who own this dataset is to develop models that predict a person's stress, happiness, and ultimately, depression. Because physical activity (steps), sedentary activity (stillness), and movement speed are all relevant to depression, the researcher must correctly extract these features from the data. Furthermore, a peak with a particular waveshape in the EDA signal represents an electrodermal response (EDR), which occurs due to increased sympathetic nervous system activity and can be indicative of increased emotion and stress; therefore EDRs must be correctly detected as well. However, EDRs that occur during movement or when the person's body temperature is high are less likely to relate to stress. Thus an understanding of the interrelations between multiple features and signals plays a central role in deducing how to best analyze the data. Rotator can be used to test flexible audio-visual representation as an alternative approach for interpreting high-dimensional data.

Rotator was originally designed for applications in which data derived from a network of sensors is analyzed. However, the dataset under discussion is drawn exclusively from a single wrist-worn sensor. Therefore, a sensible geometric layout was imposed in which each data node is spread across a human body contour line (Figure 4, left). A node map associated with a physical body enables spatialization of streams with respect to a reference map, and is thought to expedite the training process for new users.

The Biosensor dataset lends itself well to sonification in that many data types are present: transients (steps, EDR), continuous rapidly varying signals (accelerometer, EDA) continuous, slow-changing signals (body temperature), and binary signals (artifact detection). A sonification scheme designed for five features is summarized in Table 1 and an example of sound waves juxtaposed with their corresponding raw data streams is shown in Figure 5. Clicks and pulses are used for the transient data, which are particularly readily perceived when spatialized with respect to the listener. The EDA signal is sonified as white noise with a cutoff frequency modulated by the data. The user may intuitively draw a connection between the rushing-water sound of white noise and the activity level of the skin. Furthermore, it has been demonstrated in prior literature that transients and noise are most readily distinguished by the listener. Recalling discussions from Section 3, this phenomenon is largely due to the fact that transients are most likely to have distinctive onset times. Next, the accelerometer data is sonified as a high-frequency-band oscillator with modulated frequency. The rapidly changing nature of the raw accelerometer data lends itself well to this simple oscillator audification. Body temperature was sonified using a low-frequency-band beat effect synth with beat frequency corresponding to deviation from average, and wave shaping used to distinguish between positive and negative deviations. Thus, the user can determine how far body temperature has drifted, and in what direction. The frequency of the temperature synth is segregated from that of the accelerometer synth by at least 6 semitones, bearing in mind the van Noorden diagram temporal cohesion boundary for pitch segregation (see [7]). Taken together, the user is able to derive a feeling for the stress and activity levels of the person under study by growing acquainted with the interplay between variables as well as validating hypotheses by shifting data between their visual and auditory senses.

6. EVALUATION

6.1. Overview

A small-scale user study was conducted in order to gauge the influence of flexible audio-visual data display on a user's ability to

Accelerometer	Oscillator with modulated frequency
EDA	Filtered white noise with modulated cutoff
Temperature	Beat effect + wave shaping
Steps	Clicks generated using envelope function
EDR	Oscillator-driven pulse

Table 1: Biosensor Sonification Scheme Summary

Temperature	**EDA/EDR**	**Accel/Steps**	**State**
Low	High	Low	Stressed
High	High	Low	Emotional
High	Either	High	Active

Table 2: Excerpt from an information sheet provided to user study participants for characterizing data samples

draw conclusions about the structure of a dataset. In particular, the study is designed to measure a user's perceived cognitive load under different perceptual conditions, as well as measure the speed and tentative accuracy with which the user completes a task. The Biosensor dataset was chosen for use in this study because it included the most varied and developed sonification mappings of the three application areas considered in Section 5.

6.2. Methodology

The biosensor dataset consists of five parameters listed in Table 1 and drawn from one day of stressful activity and one day of calm activity. Participants were asked to classify approximately six-minute-long excerpts[2] from the time series data set on a 5-point ranking of stressed/relaxed and active/still. The users were asked to perform this task for between five and ten data samples under four perceptual conditions, which all users experienced in the following order: (1) all streams presented as audio ('A'), (2) all streams presented as plots ('V') (3) all streams presented as audio and up to one stream at a time simultaneously presented visually ('AVS') (4) A subset of streams presented as audio and up to one stream at a time presented visually; the visual stream *cannot* also be sonified ('AVD'). Conditions (3) and (4) distinguish between a scenario in which the user is observing a plot that they simultaneously hear and a scenario in which the user is seeing plots that they do not simultaneously hear. As long as the conditions of each trial were met, the users were free to move around the audio and visual boxes (see Figure 1 as a reminder) as well as adjust the synthesizer mapping parameters available in the lefthand panel. After rating each sample on the basis of stress level and activity level, the user presses a button to load the next sample. Browser localStorage was used to collect responses.

Each of five participants spent approximately two hours using the Rotator platform. Participants did not have any hearing impairments. Audio was spatialized using Web Audio's head related transfer function (HRTF) panning model and presented to the user via traditional headphones. The first ~45 minutes served as a training session. The training session consisted of the following steps:

- Verbal overview of per-stream data-to-audio mapping
- Review sample audio clips and visualizations for each stream
- Review Table 2, a qualitative guide for assessing stress and activity levels based on available data
- Review sample audio clips and visualizations for extreme states: high-stress, low-stress, high-activity, and low-activity
- Review samples of all five audio streams playing at once

If a data sample is deemed to meet the second criterion in Table 2 ('emotional'), then the participant was instructed to rate the user as having some moderately increased stress level. It is important to note that there are many biological caveats to the heuristics provided in Table 2. For the purpose of a perceptual study, it is only important for the provided identification instructions to be consistent across all user study participants.

After completing trials under each of the four perceptual conditions, users were asked to fill out a NASA task load index (NASA-TLX) survey, a widely used assessment tool for ranking perceived cognitive workload to complete a task [19]. Finally, at the end of the study, users filled out a final survey regarding their subjective experience performing the task under each condition.

6.3. Results

The rounded and weighted NASA TLX scores for each participant after completing trials in each sensory mode are provided in Table 3. We make a few key observations from the results in Table 3: firstly, the participants' self-reported prior audio experience correlates in all 5 cases with their perceived task load ranking for the audio-only trial. Note that while in all cases, NASA-TLX cognitive load measurement for the auditory-only scenario were highest, 4 out of 5 users also stated that of all the conditions, their performance in the all-audio condition seemed to improve the most (this result, a survey question, is not pictured). This result reaffirms the importance of a training period in auditory display comprehension. Furthermore, we note that users remained most comfortable and efficient in the visual-only mode. However, there were only increases in perceived task ease in the auditory modes, which also suggests that additional training may affect the 'AVS' vs. 'V' comparison.

Secondly, and most critically, we observe in Table 3 that for 4 out of 5 participants, the 'AVS' task was ranked as requiring nearly equivalent cognitive load as the all-visual task (the 'AVS' task, as a reminder to the reader, is the task in which all audio was sonified but one stream at a time could also be visualized, functioning as a 'peeking' mode). For the one participant for whom this was not the case, the 'AVS' task was still ranked as the easiest to complete among the three modes involving audio (NASA TLX of 69 compared to NASA TLX 77 for 'AVD' and NASA TLX of 83 for 'A'). Furthermore, this subject verbalized extreme lack of familiarity working with audio. The post-experiment survey provides further validation for the 'AVS' mode: 4 out of 5 users included the 'AVS' state among trials in which they felt that their performance improve over time (as compared to 3 out of 5 users indicating improvement in the 'AVD' and 'A' modes and 1 out of 5 users indicating improvement in the 'V' mode).

There are a number of confounding factors to consider. Firstly, participants are only exposed to one possible visualization methodology, so it is possible that a different visualization approach would impact results. Secondly, since the 'A' task was performed prior to the 'AVS' and 'AVD' tasks for all participants, it is possible that participants grew more accustomed to the sonification during the

[2]Note that the user could set arbitrary rates for scanning the data

Participant	TLX (‘A’)	TLX (‘V’)	TLX (‘AVS’)	TLX (‘AVD’)	Audio Experience?
1	61	51	43	n/a	A lot
2	69	42	46	52	A lot
3	73	33	33	48	None
4	58	51	50	54	Some
5	83	38	69	77	None

Table 3: Summary of NASA TLX weighted scores for each of 5 user study participants under four different perceptual modes, which self-described audio experience marked for each participant. TLX scores are out of 100, with higher TLX scores indicate greater perceived cognitive workload, and lower scores indicate less perceived cognitive workload. See body of text for descriptions of perceptual modes

	1 (easiest)	**2**	**3**	**4**	**5 (hardest)**
clicks	5	0	0	0	0
white noise	0	3	0	2	0
envelope	0	1	3	0	1
oscillator	0	1	2	2	0
beat effect	0	0	0	1	4

Table 4: Ranking of the five synthesizers on the basis of how readily identifiable they are when played in tandem. 1 is easiest to identify, 5 is hardest to identify

latter tasks. However, based on textual user reports, the visual peeking feature in the ‘AVS’ mode was particularly useful as a means of validating any auditory cues. One participant writes ‘I definitely felt the most confident in the ‘AVS’ scenario, because I could get a quick sense from the audio and then allay any questions/concerns with a few targeted visual queries.’ Another writes: ‘The visual information was especially helpful in verifying what I was hearing.’ Users in the ’AVS’ mode opted for varying number of visual validations using the sliding visual window based on the audio sample under examination. A quantitative study comparing how audio properties influence user interaction with the sliding windows could be worthwhile future work to undertake.

6.4. Evaluation of Synthesizers

Users were asked to rank each of the 5 synthesizer types on the basis of how easily identifiable they are when played in synchrony. Rankings are shown in Table 4. Despite discrepancies in prior audio experience and perceived task workload, there was great consistency among the rankings. All users rated the steps as easiest to identify (sonified using the clicks synth) and 3/5 users rated the EDA second easiest to identify (sonified using the noise synth). Noise and transients are known to be particularly readily distinguished from one another. On the other hand, temperature data, sonified using the beat synthesizer, was least readily identifiable by participants. A plausible explanation involves the additional steps required to learn the data encoding scheme used for beat synthesizer e.g. remembering what high and low beat frequencies indicate, as well as identifying the two wave shapes corresponding to above-average and below-average. Perhaps with additional training this stream would become more readily identifiable. As a qualitative note, users appeared to only occasionally adjust the frequency and gain controls for each synth and did so mainly as a mechanism for studying the sonifications, according to post-experiment surveys. Otherwise users generally preferred to keep the controls fixed and rely on the sliding windows for adjustments.

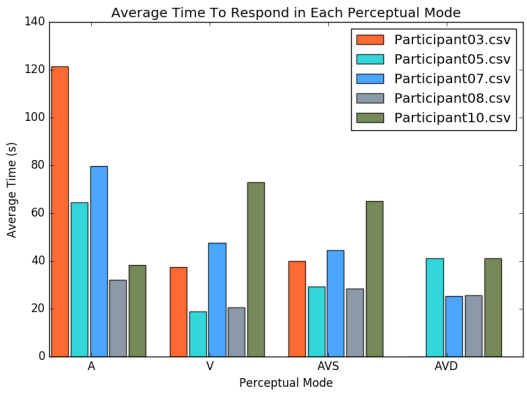


Figure 6: Ranking of average time spent in each mode for each participant. Participant 03 did not complete the ’AVD’ trials; participants 03 and 07 self-reported as having more audio experience

6.5. Evaluation of Task Completion Time

Many users requested that future studies include additional training time beyond the alotted 45 minutes, and in particular would appreciate immediate performance feedback during the training. However, there were no discernible trends in task completion time over multiple trials. Furthermore, Figure 6 shows that for most participants, the time required to complete the dual audio-visual tasks was comparable to completion time for the purely visual task. Figure 6 also shows that for participants with more experience working in audio (participants 03 and 07), auditory tasks took longer to complete despite taking less subjective cognitive effort according to the NASA-TLX. This may be due to an increased capacity for discerning subtlety in audio.

6.6. Evaluation of Task Accuracy

It is difficult to verify the accuracy with which users completed each task since there is no source of ground truth for the stress or activity level represented by the data. However, we studied various coarse accuracy measurements. For example, the number of detected steps serves as the closest coarse 1-dimensional measure for activity level. However, recalling Table 2, if users ranked purely according to step count, they would be failing to take into consideration the additional biosensor parameters. Therefore, assessments of this sort should be treated as very approximate. Figure 7 shows how users ranked activity level based on step count across all available trials for each sensory mode. The audio-only mode

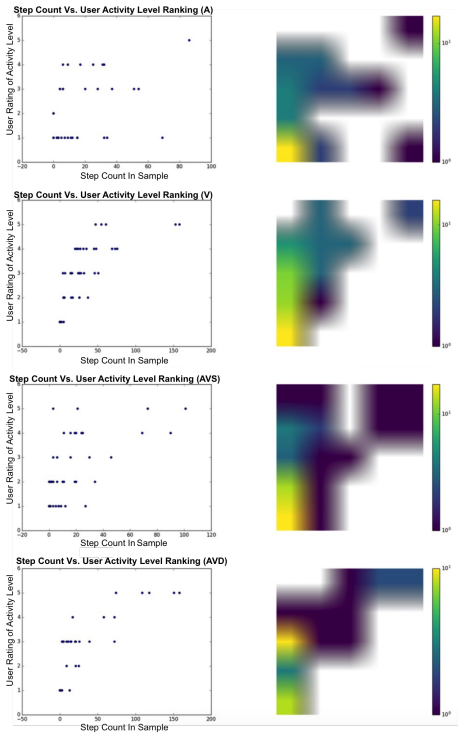


Figure 7: Scatter plot and heat map showing step count (x-axis) plotted with respect to user ranking of activity level (y-axis) for modes 'A' (top) 'V' (second) 'AVS' (third) and 'AVD' (bottom)

shows the most spread in ratings, suggesting the highest degree of perceptual ambiguity. The remaining 3 modes show very low false-positives in user perception, demonstrating the capacity for a dual audio-visual mode to improve perception.

Further, a cursory application of linear discriminant analysis across the dataset suggests that while high step count correlates most strongly with high activity rating in each of the four modes, high temperature count correlates more strongly with high activity rating only in the three modes containing auditory feedback, our first hint at improved performance in auditory modes.

7. SUMMARY AND FINAL THOUGHTS

Rotator has been used to enable a study of flexible audio-visual presentation modes for data display. One promising conclusion is the noteworthy jump in perceived task ease, task completing time, and tentatively in performance accuracy for the 'AVS' presentation mode as compared to the audio-only presentation mode. It is quite likely that optimal presentation modes are data and task dependent. Thus, platforms that provide flexibility of data stream display mode may be most broadly applicable. As a final thought, perhaps geometrically high-dimensional data can be better represented in an n-dimensional audio-visual space, with high-dimensional rotations used to adjust both sonic and visual projections.

8. REFERENCES

[1] W. Barfield, C. Hendrix, O. Bjorneseth, K. A. Kaczmarek, and W. Lotens, "Comparison of human sensory capabilities with technical specifications of virtual environment equipment," *Presence: Teleoperators & Virtual Environments*, vol. 4, no. 4, pp. 329–356, 1995.

[2] J. J. Todd, D. Fougnie, and R. Marois, "Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentional blindness," *Psychological Science*, vol. 16, no. 12, pp. 965–972, 2005.

[3] M. Czerwinski, N. Lightfoot, and R. M. Shiffrin, "Automatization and training in visual search," *The American journal of psychology*, pp. 271–315, 1992.

[4] P. Cariani and C. Micheyl, "Toward a theory of information processing in auditory cortex," in *The Human Auditory Cortex*. Springer, 2012, pp. 351–390.

[5] S. Barrass and V. Best, "Stream-based sonification diagrams," in *Proceedings of the 14th International Conference on Auditory Display*, 2008.

[6] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[7] F. Almonte, V. K. Jirsa, E. W. Large, and B. Tuller, "Integration and segregation in auditory streaming," *Physica D: Nonlinear Phenomena*, vol. 212, no. 1, pp. 137–159, 2005.

[8] E. J. e. a. Allen, "Symmetric interactions and interference between pitch and timbre," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1371–1379, 2014.

[9] H. J. Song and K. Beilharz, "Concurrent auditory stream discrimination in auditory graphing," *Journal of Computers*, vol. 3, pp. 79–87, 2007.

[10] B. N. Walker and J. T. Cothran, "Sonification sandbox: A graphical toolkit for auditory graphs," in *Proceedings of the 2003 International Conference on Auditory Display*, 2003.

[11] O. Ben-Tal, J. Berger, B. Cook, M. Daniels, and G. Scavone, "Sonart: The sonification application research toolbox," in *Proceedings of the 2002 International Conference on Auditory Displa*, 2002.

[12] S. Barrass, "Personify: a toolkit for perceptually meaningful sonification," in *ACMA*. Citeseer, 1995.

[13] I. A. Popa, J. E. Boyd, and D. Eagle, "Muse: a music-making sandbox environment for real-time collaborative play," *perspective*, vol. 3, p. 4, 2015.

[14] "Root data analysis framework," retrieved August 19th, 2016 from https://root.cern.ch/.

[15] G. Haussmann, "react-webaudio," retrieved August 19th, 2016 from https://github.com/Izzimach/react-webaudio.

[16] C. Lavor, L. Manssur, and R. Portugal, "Shor's algorithm for factoring large integers," *quant-ph/0303175*, 2003.

[17] S. Ting, "Alpha magnetic spectrometer," in *39th COSPAR Scientific Assembly*, vol. 39, 2012, p. 1975.

[18] "Snapshot study," granted data sample by Natasha Jaques and Sara Taylor. Study summary is available at http://snapshot.media.mit.edu/.

[19] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Advances in psychology*, vol. 52, pp. 139–183, 1988.

Paper Session 3
Navigation & Noise

FREQUENCY-SELECTIVE SILENCING DEVICE FOR DIGITAL FILTERING OF AUDIBLE MEDICAL ALARM SOUNDS TO ENHANCE ICU PATIENT RECOVERY

Joseph J. Schlesinger, MD

Vanderbilt University Medical Center
Department of Anesthesiology
1211 21st Ave, MAB 526
Nashville, TN, 37212
Joseph.J.Schlesinger@Vanderbilt.edu

Elizabeth Reynolds, Brittany Sweyer, Alyna Pradhan

Vanderbilt University
Department of Biomedical Engineering
2301 Vanderbilt Place
Nashville, TN, 37235
Brittany.C.Sweyer@Vanderbilt.edu
Alyna.M.Pradhan@Vanderbilt.edu
Elizabeth.L.Reynolds@Vanderbilt.edu

ABSTRACT

Free-field auditory medical alarms, although widely present in intensive care units, have created many hazards for both patients and clinicians in this environment. The harsh characteristics of the alarm noise profile combined with the frequency at which they sound throughout the ICU have created discomfort for the patients and contribute to psychological problems, such as PTSD and delirium. This frequency-selective silencing device seeks to attenuate these problems by removing the alarm sounds from the patient perspective. Patients do not need to hear these alarms as the alarms primarily serve to alert clinicians; therefore, this device, using a Raspberry Pi and digital filters, removes the alarm sounds present in the environment while transmitting other sounds to the patient without distortion. This allows for patients to hear everything occurring around them and to communicate effectively without experiencing the negative consequences of audible alarms.

1. INTRODUCTION

Significant issues plaguing successful patient recovery in Intensive Care Units (ICUs) are the frequent occurrence of clinical alarms and the harsh, shrill noises that generally characterize these sounds. Alarms sound frequently to alert clinicians of physiologic aberrancy that exceeds a threshold - yet many alarms have low positive predictive value, meaning that there are high rates of false positives indicated by alarms [1]. As stated by Edworthy, multiparameter auditory warnings can be combined to create varying degrees of urgency [2]. Although the utilization of these results has proven useful to alert clinicians of possible danger, the potential negative consequences from the piercing alarm sounds were not considered from the patient perspective. These alarms have been responsible for numerous negative consequences for both patients and physicians in the ICU. While clinicians can suffer from alarm fatigue and desensitization, the patient-specific consequences are of the utmost concern, as patients commonly experience Post Traumatic Stress Disorder (PTSD) anchored to critical illness and delirium after a stay in the ICU, as assessed by the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5) and the Confusion Assessment Method for the ICU (CAM-ICU), respectively [4]. According to Wade, 88% of the ICU patients interviewed experienced hallucinatory/delusional intrusive memories related to ICU care for up to 8 months after hospital discharge [3]. These memories were not factual but rather were fabricated memories or ideas that may have been based on or influenced by experiences in the ICU. Additionally, Pandharipande found that prior to ICU admission, only 6% of patients showed evidence of mild-to-moderate cognitive impairment [4]. After discharge from the ICU, this number increased to 25% of patients.

While the underlying causes of these disorders are not determined, the frequent, loud noises produced by clinical alarms often wake patients in the middle of the night, disturb their sleep patterns, and sound for extended lengths of time with no healthcare provider explanation to the patient regarding the reason behind the alarm. In work from Lutter, 67.2% of the alarms from three different machines in the ICU were false positives [5]. These situations can be incredibly disorienting for patients and could potentially contribute to psychological problems after discharge from the ICU. Furthermore, the fact that false alarms are so prevalent further justifies the fact that they cause an unnecessary amount of noise exposure in the ICU environment. If the alarm sounds are not detected by the patient, the likelihood of developing these psychological disorders may be reduced. It is critical to understand that patients do *not* need to hear alarms in these environments; the information conveyed by medical equipment is to signal to physicians and nurses to act, but patients themselves do not need to hear the harsh and shrill tones that occur with free-field audible medical alarms.

Future projects can completely avoid free-field alarms by transmitting signals directly to in-ear devices worn by physicians and nurses that correspond to the patients to whom they are specifically attending. As a step-wise approach, this initial patient-focused device can act as an

https://doi.org/10.21785/icad2017.062

interim solution and help alleviate the problems experienced by the patients while the more involved physician-focused device is developed.

This approach expands on the concept of avoiding unnecessary alarms in already busy and noisy hospital settings. This is accomplished by the creation of a wearable frequency-selective silencing device. The solution to this problem silences the frequencies corresponding to the alarm noises (primarily patient monitor red/crisis alarm) and will allow the passage of all normal sounds (speech and other environmental stimuli), while maintaining their quality to reduce the likelihood of delirium.

2. DEVICE DESIGN NEEDS

The wearable technology must be user-friendly and comfortable to allow for continuous patient wear, especially while the patient is asleep. Additionally, the device may reduce the occurrence of PTSD and delirium in the ICU by blocking alarm sounds while allowing the passage of all other environmental noise, such as speech and TV sounds. It is important to note that overstimulation of the auditory sense as well as a complete lack of stimulation of the auditory sense can contribute to PTSD and delirium, which is why noise-cancelling headphones and/or simple earplugs that dampen all environmental noise entirely are *not* the desired solution [4].

Keeping this in mind, the device shall not muffle or distort any normal environmental sounds as this could lead to negative consequences for the patient. This also mandates that the device process environmental noise in real-time because a noticeable delay could also contribute to psychological distress for the patient. The device should also be equipped with a detection method so that the filtering system is only activated when an alarm noise is present in the environment. By including this feature, the likelihood of unnecessarily distorting speech can be reduced and ICU-induced PTSD can further be avoided.

3. DIGITAL SIGNAL PROCESSING

To remove the alarm sound, MATLAB (MathWorks, Natick MA) Digital Signal Processing was utilized to initially implement and test our digital filters. This experimental process required multiple iterations to determine the filter metrics that successfully removed the alarm sounds. A spectral analysis was performed on a single alarm sound to obtain its frequency components. Then, an Infinite Impulse Response (IIR) Elliptic bandstop filter was created to block the frequency that specifically dominated in the spectral analysis. The width of the stopband had to be optimized so that the alarm component was completely blocked yet the effect on environmental noise was minimized. The sound file was then filtered by the newly created bandstop filter, and another spectral analysis was performed to determine the next most prominent frequency component. This led to the creation of filters targeting the common red/patient crisis alarm with the most important ones focused at 960 Hz, 1920 Hz, 2880 Hz, and 3840 Hz.

The dynamic digital filter was then generated in Simulink (MathWorks, Natick MA) using the filter specifications determined in MATLAB. The design is two-fold in that it contains both a detector and a series of filters. The detector continuously processes all incoming environmental sounds and determines the power present in the unfiltered environmental noise as compared to the power present in the filtered version. If this difference exceeds a predetermined threshold, this serves to indicate that an alarm is present in the environment. If the alarm sound is detected, the detector switches on the digital filter, and the filtered version of the noise is passed to the patient. This switching mechanism is critical to the design as it ensures that unnecessary processing and potential distortion will not occur for the patient if no alarms are sounding in the environment. Implementing the detector will further confirm that the patient will not experience ICU-induced psychological problems.

The procedure involved obtaining several recordings of the alarm sound without any background noise. These recordings were used to create and design various filters that eliminated the prevalent frequencies, which comprise the shrill and harsh sound of the alarm. Then, the original alarm sound wave file (.wav) was processed using the created filter and it was shown that the entirety of the alarm was silenced.

Furthermore, another set of recordings were obtained, and these contained the alarm sound with various environmental conditions (television playing in the background, doctors and nurses speaking, presence of pulse oximetry). These recordings were processed using the same filter mentioned above to show that the sound waves associated with the alarm were entirely removed, while maintaining the integrity and ensuring the passage of all other sounds.

4. DEVICE COMPONENTS

The hardware portion of this device continuously completes the digital filtering task during the device's operation. To do this, the Simulink code for the detector and filter has been uploaded onto a Raspberry Pi (Raspberry Pi Foundation, Cambridge UK) to allow for alarm filtration. A microphone connected to the Raspberry Pi obtains and passes the environmental sound to the digital detector. As previously mentioned, the Raspberry Pi digitally filters the predetermined alarm sound frequencies while letting the other environmental sound frequencies pass. The filtered signal is then transmitted as an output to the user through passive noise cancelling earbuds. It is important that these earbuds are noise cancelling so that the alarm sound that is present in the environment does not pass through traditional headphones and leak into the sound that is heard by the patient. Using

noise cancelling headphones ensures that only the filtered alarm sound is passed to the patient and that the goal of the design is accomplished [Figure 1].

Figure 1: Frequency-Selective Silencing Device

5. PROOF OF CONCEPT

5.1. Experimental design and testing

The testing methods are two-fold, subjective and objective – the subjective testing utilizes human participants to determine if speech intelligibility is maintained with alarm filtering, as outlined in the device design needs statement.

5.1.1. Subjective testing background

We utilized the seminal approach for speech intelligibility testing, as outlined by Lehiste *et al* [6]. Intelligibility is defined as a property of speech communication involving meaning. We utilized the consonant-nucleus-consonant (CNC) paradigm for subjective speech intelligibility testing. The CNC paradigm presents monosyllabic words to the participants and the experimenter scores each word based on the number of phonemes repeated correctly. A phoneme has little lexical meaning as an unclassified speech event – phonemes are signals, not symbols. The CNC word lists are phonemically and phonetically balanced. As the term "phonetics" is normally used in American linguistics, it concerns the physiological and acoustical properties of speech [6]. The ten CNC lists are composed of 50 words each, representing a selection of 500 words from the total of 1,263 which are refined from the Thorndike and Lorge volume after phonetic/phonemic balancing [7]. As an example, participants would be exposed to a monosyllable such as "goose." The word "goose," as all monosyllables in the CNC sets, contains three phonemes. The first phoneme, the consonant is a hard "g" (say 'guh). The second phoneme, the vowel or nucleus is a prolonged "o" (say 'ooo'). And, the third phoneme, the second consonant, has a hissing "s" sound. These phonemes were individually scored for each word. If the participant uttered "goose," the entire word is marked as correct. If the participant uttered "goo" for the word "goose," the resultant score would be ~66%, as the first and second phonemes were correct.

5.1.2. Subjective testing methods

Twenty-four (24) participants ranging 18-22 years of age participated in the subjective testing paradigm. After consent and ensuring normal or corrected to normal hearing, the participants were brought into a simulated ICU setting where they laid down on a bed with a patient monitor alarm stimulus approximately 6 feet above their heads at a 45-degree angle from their head position, similar to an ICU patient monitor location. The speech stimulus was positioned 4 feet above their head and directed towards the unilateral ear as the alarm stimulus. The speech stimulus was positioned as if another person was standing at their bedside and speaking to them. To ensure there was not confounding from the alarm stimulus simply secondary to volume, and based on recently completed work on the negative signal-to-noise ratio in the Schlesinger lab, the alarm stimulus was delivered at 70 dB and the speech stimulus was delivered at 77-79 dB as verified with an Amprobe SM-10 Class II sound level meter (Amprobe, Lynbrook, NY). The participants wore Bose QuietComfort 20i earbuds (Bose, Framingham, MA) connected to the device for stimuli exposure. To prevent the subjects from "searching" for the alarm sound in the background noise, they were not told that the device is intended to filter alarm noises but simply asked if they heard alarms throughout their time during the testing and to repeat the CNC word back to the experimenter. The participants were exposed to two randomly selected sets of 50 CNC words. One set would have alarm filtering, and the other set would have no alarm filtering. The sets of 50 words were purposely different to avoid a learning effect. The study lasted approximately 15 minutes per participant with breaks offered between CNC word sets [Figure 2].

5.1.3. Objective testing methods

Objectively proving that the device accomplished the project aims, an experiment was performed to demonstrate that the frequency components specific to the alarms were missing from the filtered sound. In the initial stages of the project, a Fast Fourier Transform (FFT) was performed using MATLAB and Simulink on the unfiltered alarm sound sample and the filtered alarm to compare the magnitudes of the frequency components present between the two sounds [Figures 3a, 3b, 4a and 4b].

5.2. Results

The subjective CNC testing yielded clinically and statistically significant improvement with alarm filtering. The phoneme score improved from 42.54% (95% CI: 38.96, 46.12) to 56.71% (95% CI: 53.32, 60.10) correct with alarm filtering ($p<0.001$). The word score improved from 18.42% (95% CI: 14.62, 22.22) to 27.42% (95% CI: 23.17, 31.67) correct with alarm filtering ($p<0.01$) [Figure 2]. Additionally, besides these data, the participants endorsed a high-stress state during alarm exposure.

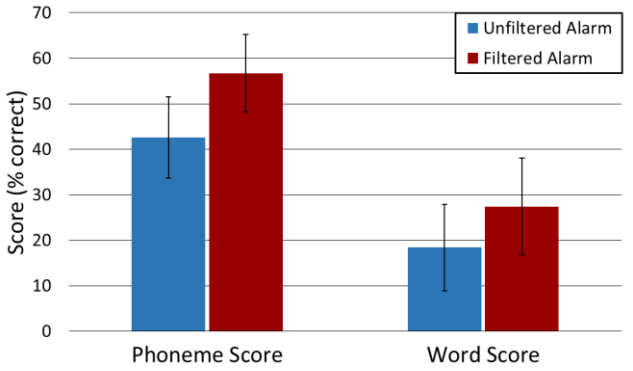


Figure 2: Improved phoneme and word score with alarm filtering (error bars signify standard deviation)

In the objective testing using MATLAB, the series of bandstop filters created on MATLAB dampened the magnitudes of the frequencies present in the alarm on the *order of 10³*, as seen in Figures 3a and 3b (note the Y-axis values).

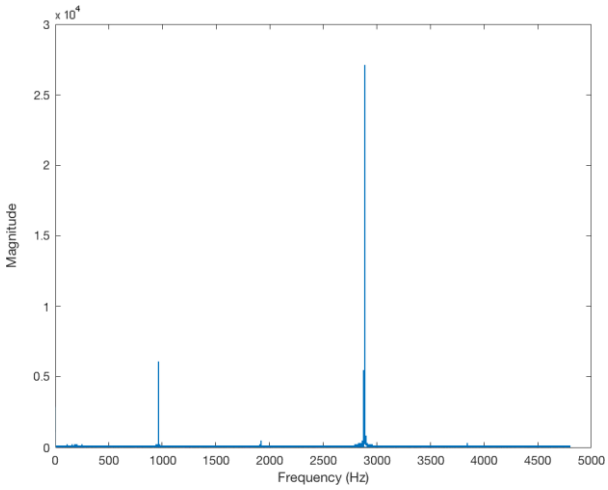


Figure 3a: FFT of a single unfiltered alarm (a) and the same alarm filtered by a series of bandstop filters (b)

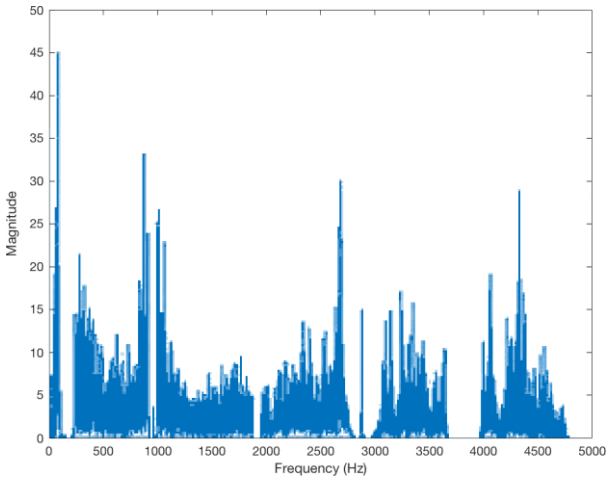


Figure 3b: FFT of a single unfiltered alarm (a) and the same alarm filtered by a series of bandstop filters (b)

Once the filtering on MATLAB proved successful, Simulink (MathWorks) was used to compile the software and deploy the data onto a Raspberry Pi device. By utilizing a file (.wav) with both the alarm sound and environmental noise present, it was proven that the Simulink software could successfully filter the alarm frequencies as shown in Figures 4a and 4b (note the Y-axis values).

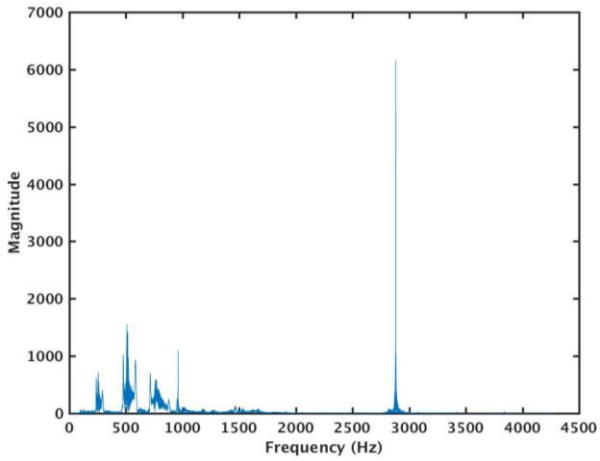


Figure 4a: Fast Fourier Transform of unfiltered real-time audio output containing speech and alarm sounds

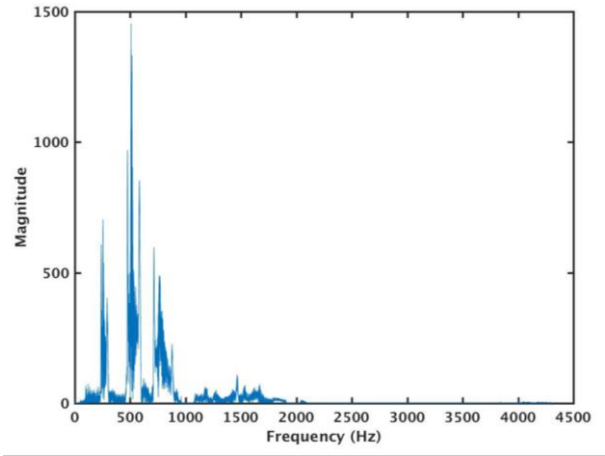


Figure 4b: Fast Fourier Transform of real-time audio output with speech and filtered alarm sounds

5.3. Expected patient benefit

Following the successful implementation of the digital signal processing onto the Raspberry Pi hardware, it is expected that the Raspberry Pi will output an audio signal that will contain the original input audio signal without the alarm-sound frequencies. The Raspberry Pi should be able to do real time filtering of audio signals input through the microphone attachment to the Raspberry Pi. Using the switch, the user should be able to hear an output signal of the filtered environmental noise only when the alarm sound is present. Furthermore, the patient should be able to hear an output signal of the original environmental noise when the alarm sound is not present; therefore, not experiencing any distortion of the sound and the only change being the elimination of the alarm frequencies.

5.4. Expected clinician benefit

While this device was designed for the patient, the future directions indicate developments for the clinician. However, with this patient-centered device, the benefits are manifold. Besides filtering out unnecessary alarms for patients, speech intelligibility will be improved between patients and the healthcare team. This is imperative during patient and family centered rounds in the ICU where complex care plans are discussed and the patient must fully understand the risks and benefits of all treatment options before giving consent. Additionally, improving patient satisfaction and minimizing disruptions in the healthcare setting by attenuating the alarm exposure for the patient may further enhance the patient-clinician relationship. In the face of decreased deleterious neuropsychological outcomes for patients, there may be decreased length-of-stay, and improved healthcare economics.

6. LIMITATIONS AND FUTURE DIRECTIONS

This device relies on the use of noise-cancelling headphones to transmit the filtered sound to the patient. Future design will incorporate a wireless, in-ear device that can perform all the necessary filtering functions and transmission of the filtered sound in the device itself. With respect to the overarching problem of audible medical alarms, this device serves as an interim solution that may solve the patient problems of PTSD, delirium and general patient discomfort. However, the physician-related problems still remain. Thus, as previously discussed, a second in-ear device is in the process of being developed that transmits alarms and patient information directly from the patient monitors and equipment in the patient room to the physician or nurse at the optimal signal-to-noise ratio. This device would suppress the need to have audible free-field medical alarms. The patient specific device described herein would remain necessary in the likely slow transition to, or absence of complete global adoption of healthcare provider in-ear monitoring devices.

7. CONCLUSION

Audible medical alarms are the cause of myriad hazards in hospital and ICU settings. Their shrill acoustic features and the frequency at which they alarm (both in sheer number and frequency spectrum) are responsible for many negative consequences, especially for patients. Patients can experience PTSD and delirium secondary to sleep disturbance from alarms and healthcare providers' divided and diminished attentional resources allocated to alarms. This frequency-selective silencing device was created to alleviate these problems and create a more comfortable environment for the patients during their length of stay in the ICU. This device has been demonstrated to successfully remove alarm sounds while avoiding audible distortion of speech and other environmental noise, and should it be widely implemented in hospital setting, it will prevent patients from hearing the disturbing and potentially harmful sounds of free-field medical alarms and may improve patient safety.

8. ACKNOWLEDGMENT

We would like to thank the Department of Anesthesiology (especially the Benjamin Howard Robbins Research Scholar Program) and the Department of Hearing and Speech Sciences at Vanderbilt University Medical Center, particularly Drs. Ben Hornsby and Matthew Weinger. We would also like to thank the Department of Electrical Engineering and Computer Science at Vanderbilt University, especially Dr. A. B. Bonds, Dr. Dean Wilkes and Garrett Hoffman. We would also like to acknowledge Dr. Matthew Walker, III, and the Department of Biomedical Engineering at Vanderbilt University, and especially Dr. Michael King for departmental support for this research.

9. REFERENCES

[1] C. Meredith, J. Edworthy, “Are there too many alarms in the intensive care unit? An overview of the problems.” *Journal of advanced nursing*, vol. 21, no.1, pp. 15-20, January 1995.

[2] J. Edworthy, S. Loxley, I. Dennis. "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency." *Human factors*, vol. 33, no. 2, pp. 205-231, April 1991.

[3] D. Wade, C. Brewin, D. Howell, E. White, M. Mythen, J. Weinman. “Intrusive memories of hallucinations and delusions in traumatized intensive care patients: an interview study.” *British Journal of Health Psychology*, vol. 20, no. 3, pp. 613-631, June 2014.

[4] P. Pandharipande, T. Girard, J. Jackson, et al. "Long-term cognitive impairment after critical illness." *New England Journal of Medicine*, vol. 369, no. 14, pp. 1306-1316, October 2013.

[5] N. Lutter, S. Urankar, S. Kroeber. “False alarm rates of three third-generation pulse oximeters in PACU, ICU and IABP patients.” *Anesthesia and analgesia*, vol. 94, no. 1, pp. 69-75, November 2002.

[6] I. Lehiste, G. Peterson. “Linguistic considerations in the study of speech intelligibility.” *The Journal of the Acoustical Society of America.* vol. 31, no. 3, pp. 280-286. March 1959.

[7] E. Thorndike, I. Lorge. “The Teacher’s Word Book of 30,000 Words. (Teachers College, Columbia University Press, New York, 1952).

INTRODUCING MULTIMODAL SLIDING INDEX: QUALITATIVE FEEDBACK, PERCEIVED WORKLOAD, AND DRIVING PERFORMANCE WITH AN AUDITORY ENHANCED MENU NAVIGATION METHOD

Ruta R. Sardesai

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
rrsardesai@gmail.com

Thomas M. Gable

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
thomas.gable@gatech.edu

Bruce N. Walker

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
bruce.walker@psych.gatech.edu

ABSTRACT

Using auditory menus on a mobile device has been studied in depth with standard flicking, as well as wheeling and tapping interactions. Here, we introduce and evaluate a new type of interaction with auditory menus, intended to speed up movement through a list. This multimodal "sliding index" was compared to use of the standard flicking interaction on a phone, while the user was also engaged in a driving task. The sliding index was found to require less mental workload than flicking. What's more, the way participants used the sliding index technique modulated their preferences, including their reactions to the presence of audio cues. Follow-on work should study how sliding index use evolves with practice.

1. INTRODUCTION

Distracted driving has long been discussed as a common cause of automobile accidents. Emerging technologies have led to advances in the availability of information in previously "disconnected" locations, for example, information presented on handheld devices as well as within vehicles in the form of in-vehicle technologies (IVTs). These IVTs include both personal multimedia such as audio, video and images as well as driver-relevant information such as navigation, weather and traffic alerts, etc. Many of these IVTs as well as hand-held devices like cellphones are interacted with through list-based menus. In traditional interactions, the menu is accessed using a touch-based approach, requiring drivers to take their eyes off the road. Thus, interacting with IVTs adds a visually demanding process on top of the already visually loaded primary task of driving. The driver inattention created as a result is a cause for growing concern. In addition, while speech based systems are becoming more prevalent, users sometimes still require a recognition-based interface such as a menu system rather than a recall-based system used in voice command interfaces.

This paper presents research investigating the use of a novel multimodal "sliding index" interface, for use in navigating lists on a mobile device. In particular the novel interface is compared to traditional flicking interactions. By looking at participants' qualitative feedback and perceived workload across four interfaces, the research aims to determine if and how the multimodal sliding index facilitates list navigation during multitasking, such as driving while using a phone.

Nees and Walker [1] point out that multimodal interaction might help reduce the overall impact of user inattention in multitasking interactions. According to multiple resources theory, spreading the modality of interaction across different senses helps users access "separate pools of modality resources" rather than over-burdening just a single pool, and results in an increase in performance [2]. By using various auditory cues, thus reducing the dependency on visual resources, performance has improved with multimodal interfaces [3,5]. In addition, enhancing basic text-to-speech (TTS) with advanced auditory cues (AACs) can result in lower workload than TTS alone [5], with participants showing a clear preference for menus with AACs [6].

In order to enhance TTS menus, research has been done particularly in the area of non-speech cues. These cues can be categorized as "menu *item* cues" and "menu *structure* cues." Item cues provide extra information about the characteristics (e.g., available vs. unavailable) of a given menu item. Spearcons, which are short sounds consisting of a sped-up version of a spoken phrase, are examples of item cues [5,6]. Item cues are particularly useful if the user (driver) is familiar with the contents of the menu. In contrast, structure cues provide enhanced information about where the user is in the menu, including concepts like scroll bars and spindex. The spindex (speech index) cue [6,7,8] is a set of brief sounds that correspond to each menu item, usually comprised of the sound of its first letter [7], thereby giving the user an overview of their location in the list. Both item and structure cue types can work well together, and with TTS. Such cues afford participants faster search times and lower subjective workloads [5,8]. Research has shown Spindex+TTS can decrease visual time off a primary task while driving, as compared to a visual-only condition [3,9].

1.1. Sliding Index

Spindex cues alone, however, may not enhance TTS auditory menus enough to allow for safe use of in-vehicle technologies (IVTs) within the driving context. Hierarchical menus, for example, may benefit from the addition of another interaction approach, such as a letter-based *sliding index* used in alphabetically sorted list-based menus [10]. This method, featured on the iPhone, allows for a coarser granularity of navigation than simply scrolling or flicking

https://doi.org/10.21785/icad2017.029

through the list. Users can jump to particular sections of a list by selecting the corresponding letter index. This index is implemented on the right border of the screen and becomes active when touched. By adding audio cues to the sliding index, creating a multimodal sliding index, it may decrease the negative impact of mobile device use while driving.

1.2. Current Study

The current study examined the impact of adding spindex+TTS audio cues to a sliding index, to enhance the use of auditory menus, particularly in the case of IVTs. Participants were asked to find a song from a list on a mobile phone, while driving in a low fidelity simulation. The menu search task was completed during four blocks of driving, using a different version of the phone interface in each block. This was followed by a semi-structured interview aimed at understanding user preferences towards the interfaces, use of technology in cars, and potential solutions to issues they faced in this space. In the study, we expected the combination of Spindex+TTS cues with a sliding index to be equal to or better than a Visuals-Only condition with regular flicking, as measured by perceived workload and preference.

2. METHOD

2.1. Participants

A total of 23 participants (18 male) with a mean age of 19.8 years and 3.5 years of driving experience took part in the research. All were required to have normal or corrected to normal vision and hearing, and a valid driver's license.

2.2. Apparatus

2.2.1. Mobile phones

Two different Android phone interfaces were used in the current study (see Figure 1). Both had the same alphabetical menu structure and the same list of popular songs from 2009. The flicking application ran on a Nexus One phone running Android 2.1 using Eclipse and was "flicked" for interaction. The sliding index interface was implemented on a MotoX phone running Android 4.4.4. Two phones were used due to difficulties in getting one to do what was required for both interactions but were considered functionally equivalent to each other, so as not to affect the results. In addition, the sliding index application included the alphabetical index at the side, which jumps to the corresponding sections of the list by tapping a letter in the index and the auditory cue of reading out the letter via spindex. Both apps were set so that using the sliding index or flicking would interrupt any currently ongoing audio. Alongside this, the lag between the selection of a letter in the sliding index and the actual jump to that section was minimized to facilitate scrubbing.

2.2.2. Phone interaction

At the beginning of each trial, the phone would speak the title of the song from the list that the participant was supposed to find. Using the cues—either the spindex sound of every song name scrolled through, or no sound—the user would move through the list in order to find that particular song. Once found, the user would tap the selected song. The interaction (user input) consisted of two types: Flicking and Sliding Index. The two display types were Visual-Only (i.e., no sound) and Spindex+TTS. Participants held the phones in their preferred hand throughout the study with their arms on an armrest to ensure the same location for each condition.

Performance on the secondary task was measured by time required to select the targeted item, as well as accuracy. The effect of the two interaction types and two sound types, as well as their combination, was measured.

2.2.3. Driving simulation

We used the Open DS "three car platoon task" in which drivers follow a lead car at a set distance, and the lead car slows down or speeds up at intervals, forcing the driver to do the same. The task was performed on a 40" LED monitor and controlled via a Logitech Driving Force wheel and pedals. The performance of the primary driving task was measured using lateral and longitudinal deviation along with brake response time.

2.2.4. Visual Behaviors

Eye glance behavior – defined as per SAE as the percent time spent by participants looking at the screen out of the total driving time [13] – was monitored with a FaceLab 5 contact free eye tracker.

2.2.5. Perceived mental workload

Perceived cognitive load was measured subjectively using the NASA Task Load Index (TLX) [4], collected after each condition. During the TLX, participants were asked to rate how much mental, physical, temporal, etc. workload they felt was caused by the tasks they had just performed on the driving simulation.

2.3. Procedure

Following training, participants completed a baseline-driving task with no other tasks. Then, participants performed each of the four combinations of Flicking or Sliding Index, and Visual-Only or Spindex+TTS in a randomized order. In each condition, participants drove for approximately seven minutes. The secondary task, during this time, was to navigate through a list of song names to find the target song using the interaction and auditory cues that corresponded to the ongoing condition. They were asked to devote 80% of their mental resources to the driving and the remaining 20% on doing the secondary phone task as quickly and accurately as possible. After the four conditions, there was a semi-structured interview intended to understand user preferences and to explore possible future work. This included current usage of technology in vehicles; which conditions they found

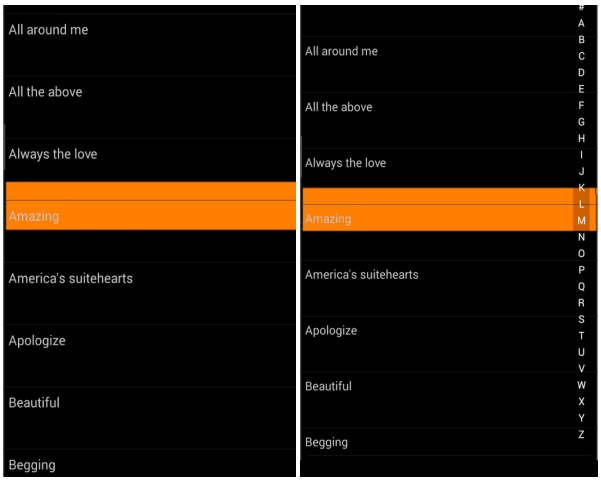


Figure 1: The two interfaces seen next to each other, with the flicking only interface on the left and sliding index system on the right.

the most and least preferable, taxing, or annoying; and a discussion about potential solutions to issues they had had with the interface.

2.3.1. Study design and analysis

The present study was a fully within-subjects 2x2 full factorial design. The factors were interaction type (sliding index vs. flicking only) and display type (visuals only vs. visual plus spindex).

Phone Accuracy (%)	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	92.17	9.98	92.17	6.00	92.17	7.99
Spindex-TTS	92.61	9.15	95.22	5.93	93.91	7.54
Interaction Type Mean	92.39	9.57	93.70	5.96	-	-
Phone Time (ms)	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	38390	4991	39838	5651	39114	5321
Spindex-TTS	40455	5184	40557	5110	40506	5147
Interaction Type Mean	39423	5087	40198	5381	-	-

Table 1. Phone performance values for accuracy (%) and average time (ms) to complete the task.

Longitudinal Deviation (m)	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	2.89	2.52	2.98	2.82	2.93	2.67
Spindex-TTS	3.03	2.52	2.48	2.02	2.76	2.27
Interaction Type Mean	2.96	2.52	2.73	2.42	-	-
Lateral Deviation (m)	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	0.27	0.16	0.31	0.17	0.29	0.17
Spindex-TTS	0.32	0.20	0.31	0.17	0.32	0.18
Interaction Type Mean	0.30	0.18	0.31	0.17	-	-

Table 2. Average values for longitudinal and lateral deviation (in meters) for the driving task.

Eyes on Road (% time)	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	73.27	26.82	74.99	25.28	74.13	26.05
Spindex-TTS	69.25	27.85	73.05	27.66	71.15	27.76
Interaction Type Mean	71.26	27.34	74.02	26.47	-	-

Table 3. Percent time eyes on the road for the four conditions.

To determine differences within each measure between the conditions, a 2x2 full factorial within-subjects analysis was conducted for interaction type and display type. Qualitative data collected through transcription during the interview were collated and entered into a database. These data were reduced by filtering only the significant content and emerging themes that were relevant to the context of the research goals [11]. Finally, thematic analysis was conducted on the remaining data in order to find emerging themes, particularly regarding potential solutions [12].

3. RESULTS

3.1. Cell Phone Performance

Values for cell phone performance can be seen in Table 1. In regards to accuracy there were no significant differences for interaction type $F(1,22)=0.66$, p=.426, sound type $F(1,22)=1.11$, $p=.304$, or any interactions $F(1,22)=0.95$, $p=.342$. There were also no significant differences in completion time for interaction type $F(1,22)=0.57$, $p=.458$, sound type $F(1,22)=2.10$, $p=.162$, or any interactions $F(1,22)=0.70$, $p=.411$. These results point to no differences between the conditions, meaning the index had no significant decrement on cell phone performance.

3.2. Driving Performance

Table 2 displays the driving task measures for the current study. The analyses found no significant differences in longitudinal deviation for interaction type $F(1,22)=2.41$, $p=.135$, sound type $F(1,22)=0.33$, $p=.571$, or any interaction $F(1,22)=4.14$, $p=.054$. There were also no significant differences found in lateral deviation for interaction type $F(1,22)=1.38$, $p=.251$, sound type $F(1,22)=2.14$, $p=.158$, or any interaction $F(1,22)=1.74$, $p=.201$. Although there were no statistically significant differences in regards to driving there was an interaction in longitudinal deviation that was approaching significance. These results point to a potential interaction of sound type and interaction type, with spindex index being the lowest deviation and swipe spindex being the largest, potentially driving this interaction.

3.3. Eyes-on-Road Time

There were no significant differences found for eyes-on-road time for interaction type $F(1,22)=2.02$, $p=.170$, sound type $F(1,22)=3.34$, $p=.081$, or any interactions $F(1,22)=0.28$, $p=.604$. This means no significant differences were found between the two interaction or stimuli types for visual time eyes on the road, the values of which are seen in Table 3.

3.4. Workload

For the TLX mental workload subscale there was a significant difference in interaction type $F(1,22)=7.45$, $p=.012$, with flicking having higher workload than sliding index. No significant difference for display type $F(1,22)=0.15$, $p=.703$, nor any interactions $F(1,22)=0.04$, $p=.852$, were found for mental workload. For the physical component of workload there was a significant difference in the display type $F(1,22)=4.93$, $p=.037$, with Spindex conditions being rated as having higher physical workload, but no significant difference was present between the interaction types $F(1,22)=4.03$, $p=.057$, nor were there any interactions $F(1,22)=0.07$, $p=.794$. There were no differences in the time/temporal portion of workload for interaction type $F(1,22)=0.01$, $p=.967$, nor display type $F(1,22)=0.20$, $p=.657$, but there was a significant interaction $F(1,22)=4.61$,

p=.043. This interaction was investigated using paired t-test post hoc analyses with Bonferroni corrections (decreasing alpha to .0083) but no significant differences were found. For total TLX workload there were no significant differences for interaction type $F(1,22)=2.18$, p=.154, display type $F(1,22)=0.29$, p=.595,, nor any interactions $F(1,22)=0.18$, p=.678. The data from the TLX scores can be seen in Table 4.

These results suggest the act of flicking results in greater perceived mental workload than sliding index conditions. In addition, conditions with Spindex+TTS resulted in higher perceived physical workload than Visual-Only conditions.

Total Workload	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	52.28	20.66	47.83	19.82	50.05	20.24
Spindex-TTS	52.72	20.53	49.93	18.66	51.33	19.59
Interaction Type Mean	52.50	20.59	48.88	19.24	-	-
Mental Workload	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	55.43	26.11	48.70	26.21	52.07	26.16
Spindex-TTS	55.87	24.62	50.00	23.65	52.93	24.13
Interaction Type Mean	55.65	25.36	49.35	24.93	-	-
Physical Workload	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	38.04	23.20	32.39	20.11	35.22	21.65
Spindex-TTS	42.61	25.89	35.87	22.60	39.24	24.24
Interaction Type Mean	40.33	24.54	34.13	21.35	-	-
Temporal Workload	**Swipe**		**Index**		Sound Type Mean	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual-Only	45.00	21.58	39.13	27.04	42.07	24.31
Spindex-TTS	37.61	23.64	43.26	26.48	40.43	25.06
Interaction Type Mean	41.30	22.61	41.20	26.76	-	-

Table 4. Workload values across the 4 conditions for total workload and the significant subscales.

3.5. Qualitative Thematic Analysis

A number of themes were found from user interviews that followed the completion of the four conditions.

3.5.1. Effortful flicking

Flicking, particularly without the added benefits of auditory cues like Spindex+TTS, was found to be tiresome, with several participants commenting that whereas they felt they had greater control navigating via swiping, they felt the *"…need to get to the general area without swiping so much," or that the "…longer scroll makes (them) more panicky."*

3.5.2. Issues with Sliding Index

Participants expressed the fact that using the sliding index had greater penalties than swiping, as they would *"…keep missing the letter and get to a whole different section*." However, they also clearly believed that the use of the sliding index, either with or without auditory cues, helped them *"…pay more attention to the road."*

3.5.3. Issues with sound

It was interesting that a few participants *"ignored the sound"* as they felt that using auditory cues *"…takes more time, but less attention,"* a sentiment that echoes Ranney, et al.'s findings regarding a driver's willingness to engage [13]. In addition, participants who felt the auditory cues were *"…too distracting"* believed they would be more open to them *"…if the voice or pace were different."* A few were supportive of the condition with sliding index and Spindex+TTS, without the use of TTS, saying *"the repeated names were annoying."*

3.5.4. Familiarity with auditory cues

There was, of course, a general lack of familiarity with auditory cues. Participants were not *"…expecting index with sound"* and believed that *"[sliding] index plus sound threw (them) off."* Some felt that although "*...sound would be helpful if I got used to it*," they "*…wouldn't get used to it without using it outside driving.*"

3.6. User Preferences

User preferences for most annoying, least annoying, most attention paid to the road, least attention paid to the road and overall preference frequencies were measured and are reported in Table 5. While participants preferred using the sliding index as an interaction method, preferences towards the presence or absence of auditory cues was not as clear. This was interesting when coupled with the fact that participants believed they had the most attention on the road using the sliding index with Spindex+TTS auditory cues. Flicking as an interaction method was rated as being more distracting, though this was not reflected in quantitative data. Participants also found the use of Flicking combined with Spindex+TTS cues to be the most annoying, while sliding index with Visual-Only was reported as least annoying.

Preferences	**Spindex Index**	**Visuals Index**	**Spindex Flicking**	**Visuals Flicking**
Overall preference	9	7	4	1
Most annoying	2	0	12	7
Least annoying	7	10	2	3
Most attention to road	9	6	5	1
Least attention to road	3	5	8	7

Table 5: Preference rating averages across participants following their completion of the study with the 4 conditions from the post-study survey.

4. DISCUSSION

The present study found that although there were no significant differences between the four conditions in cell phone performance, driving performance, or time with drivers' eyes on the road, there were significant differences in mental workload between the conditions and some interesting qualitative results. This lack of differences in most of the quantitative measures is not necessarily a bad thing, as it showed that the novel auditory interface was no worse than the currently used interfaces. It may be that with more practice these results could change as participants

become more familiar with the cues; that has been seen in previous work with these types of cues [8]. In addition, although not meeting the threshold for significance, the result found for longitudinal deviation of a potential interaction being present is nothing to ignore as it points to drivers with the spindex index being potential safer drivers.

One of the major takeaways from the current study that did meet the threshold for statistical significance was the lower mental workload for the sliding index interaction method while driving, as seen through the NASA-TLX data. This continues to point to issues with flicking as an interaction method, and while it may be the norm for lists on most hand-held touch screen devices, its use in driving scenarios is less than ideal, even when visual workload is not affected. In the qualitative analysis, the flicking method led to the most annoyance and distraction in general. When this was discussed with participants, they mentioned voice control as a substitute, though they clarified that it would work better for finding a specific object than browsing a list. Another suggestion was the use of a single touch method of interaction that would auto-scroll without requiring additional hand movements. This interaction is also known as a "push" menu [14].

Participants also felt strongly about "penalties" associated with using the sliding index. If the wrong letter were selected, the list would jump to a completely different section. This was jarring when contrasted with the easy recovery from error while flicking – a simple motion in the opposite direction. To prevent this, the letters in the sliding index need to be spread farther apart to increase the error margin and provide greater control.

Participants reported via the TLX data that the use of audio cues was physically taxing. In particular, spindex+TTS was found to be especially tiresome in combination with the sliding index interaction. In the qualitative data some expressed a concern that auditory cues could distract from music playing in their cars and exhorted that using them be optional. It would be important to determine the type of sounds that would be most preferred and least disruptive in applying these types of displays in the real world.

As mentioned, participants were unfamiliar with auditory cues and extremely familiar with flicking. Many of their comments revolved around preferring to use auditory cues with the sliding index only after becoming accustomed to it. Since the participants would only use auditory cues while driving, they said they would unlikely reach the skill levels required to commit fully to using auditory cues.

This raises a common design challenge: balancing what consumers want – either no or optional auditory cues – with what they may potentially need – mandatory use of auditory cues in order to facilitate the initial learning. Learnability is a major component of the usability of a system [15], especially in this case, as the auditory cues are transient and unfamiliar. In this respect, Spindex cues, with the benefit of pre-existing, natural mapping to their corresponding menu items, may be learned more easily [6]. When performance improvements were monitored across time, participants seemed to continue to learn and develop skills, with additional practice [5]. As such, a longitudinal study of the impact of learning on performance and ease of use of auditory cues could provide additional understanding of these cues.

4.1.1. Potential redesign

Interviews showed a need to scroll without effort. Fig. 2 shows one option: auto-scroll using a single touch. This could expand on a current iPhone feature in which tapping at the top of the screen moves the window to its top-most position. Similar functionality could allow a tap at the bottom of the screen to move a window towards its bottom-most position. Scroll rate could be controlled by the iPhone's Force Touch capability. In other phones, moving the interacting finger on the screen could control movement rate.

5. CONCLUSIONS

We compared the use of enhanced auditory cues to visual-only systems, for interacting via flicking and sliding index systems on a menu item selection task while driving. Subjective workload differences from the NASA-TLX suggest that the sliding index was better than flicking, but the auditory cues created additional perceived physical workload. However, the qualitative results revealed a large amount of interesting data such as greater effort involved in flicking, and participants thinking they would be better at the task with more practice. It was these data points from the qualitative data that was used to create a new ideation of the interaction method, helping to further improve the system.

The results of this research suggest that this new method of interacting with auditory menus for list-based systems is highly influenced by previous practice. The participants' familiarity with flicking seemed to be more of an issue than the use of the auditory menus. It may be that more practice would give participants the experience needed to perform that form of interaction more efficiently and to use the auditory cues more efficiently due to the slow nature of the auditory displays that participants reported. It may however be that the use of this interaction method when paired with auditory displays has too high of a cost to miss the target, and should be avoided or highly trained before use.

Figure 2: Mock-up of a potential re-design to improve the interface including the use of force touch to control the rate of movement down the list.

6. ACKNOWLEDGMENTS

Portions of the work were supported by a National Science Foundation (NSF) Graduate Research Fellowship (DGE-1148903), as well as additional grant funding from the NSF and from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR).

7. REFERENCES

[1] Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of human factors and ergonomics*, *7*(1), 58-99.

[2] Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449-455.

[3] Gable, T. M., Walker, B. N., Moses, H. R., & Chitloor, R. D. (2013). Advanced auditory cues on mobile phones help keep drivers' eyes on the road. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 66-73). ACM.

[4] Hart, S.G., & Staveland, L.E. 1988. Development of NASA- TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, by Meshkati, N., & Hancock, P.A., 239-250. Amsterdam: North Holland Press.

[5] Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu navigation with in-vehicle technologies: Auditory menu cues improve dual task performance, preference, and workload. *International Journal of Human-Computer Interaction*, *31*(1), 1-16.

[6] Jeon, M., & Walker, B. N. (2011). Spindex (speech index) improves auditory menu acceptance and navigation performance. *ACM Transactions on Accessible Computing (TACCESS)*, *3*(3), 10.

[7] Jeon, M., & Walker, B. N. (2009). "Spindex": Accelerated Initial Speech Sounds Improve Navigation Performance in Auditory Menus. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 17, pp. 1081-1085). SAGE Publications.

[8] Jeon, M., Walker, B. N., & Srivastava, A. (2012). "Spindex" (Speech Index) Enhances Menus on Touch Screen Devices with Tapping, Wheeling, and Flicking. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *19*(2), 14.

[9] Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

[10] Sun, Q., & Hurst, W. (2008). Video Browsing on Handheld Devices&# x2014; Interface Designs for the Next Generation of Mobile Video Players. *MultiMedia, IEEE*, *15*(3), 76-83.

[11] Krathwohl, D. R. (1998). Methods of educational and social science research. Long Grove, IL.

[12] Taylor-Powell, E., & Renner, M. (2003). Analyzing qualitative data.

[13] Ranney, T. A., Mazzae, E., Garrott, R., & Goodman, M. J. (2000). NHTSA driver distraction research: Past, present, and future. In *Driver distraction internet forum* (Vol. 2000).

[14] Yalla, P., & Walker, B. N. (2007). *Advanced Auditory Menus*. Georgia Institute of Technology GVU Center Technical Report # GIT-GVU-07-12. October.

[15] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-computer interaction*. Addison-Wesley Longman Ltd

AUDITORY AND HEAD-UP DISPLAYS FOR ECO-DRIVING INTERFACES

Woodbury Shortridge

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
whshortridge@gmail.com

Thomas M. Gable

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
thomas.gable@gatech.edu

Brittany E. Noah

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
brittany.noah@gatech.edu

Bruce N. Walker

Sonification Lab,
Georgia Institute of Technology,
654 Cherry Street, Atlanta, GA, 30332, USA
bruce.walker@psych.gatech.edu

ABSTRACT

Eco-driving describes a strategy for operating a vehicle in a fuel-efficient manner. Current research shows that visual eco-driving interfaces can reduce fuel consumption by shaping motorists' driving behavior but may hinder safe driving performance. The present study aimed to generate insights and direction for design iterations of *auditory* eco-driving displays and a potential matching head-up visual display to minimize the negative effects of using purely visual head-down eco-driving displays. Experiment 1 used a sound card-sorting task to establish mapping, scaling, and polarity of acoustic parameters for auditory eco-driving interfaces. Surveys following each sorting task determined preferences for the auditory display types. Experiment 2 was a sorting task to investigate design parameters of visual icons that are to be paired with these auditory displays. Surveys following each task revealed preferences for the displays. The results facilitated the design of intuitive interface prototypes for an auditory and matching head-up eco-driving display that can be compared to each other.

1. INTRODUCTION

From 1990 to 2007 transportation has been responsible for a 45% growth in CO_2 emissions, with a predicted rise of an additional 40% by 2030 [1]. Emerging innovations in vehicles are aimed at improving fuel economy (FE) to reduce emissions and reduce cost of ownership. Saving fuel can immediately reduce cost of operation and environmental impacts. *Eco-driving* is a readily available technique that shapes driving behaviors increase FE without reliance on automotive advances such as body or engine changes. Research shows that driving styles such as rapid acceleration and deceleration hinder eco-driving; and therefore, are used as prompts for eco-driving displays [2].

https://doi.org/10.21785/icad2017.028

Herein we discuss efforts to develop low workload displays for eco-driving, notably, through auditory displays and visual head-up displays.

2. Current Eco-Driving Interfaces

Fuel Economy Driver Interfaces (FEDIs) have been shown to improve FE by up to 20% [3]. However, nearly all research and development of FEDIs has focused on visual displays [4], with most being head-down, dashboard displays. Figure 1 shows the TOYOTA Eco-Indicator, an eco-indicator bar that tells you how economically you are driving. When the driver is accelerating excessively, the bar will stretch beyond the "eco zone" and start flashing. That means it is likely that more fuel than needed is being used. Unfortunately, driving is a demanding task when it comes to visual attention [5] and since most current FEDIs also rely on visual resources, this may create a competition for additional resources. Evidence supports the case that current visual FEDIs can distract drivers from attending to the road, increase workload, and in effect, hinder driving performance [4].

2.1. Design Considerations

While current dashboard-based FEDIs may increase distraction, there are other approaches to design eco-driving interfaces that limit driver distraction while simultaneously shaping motorists' driving behaviors. In order to increase driver safety and FE, head-up displays (HUDs) and auditory displays should be explored.

Figure 1: The TOYOTA Eco-Indicator: A visual-only dashboard display of fuel economy.

2.1.1. Auditory Displays

Wickens' Multiple Resource Theory (MRT) [6] provides valuable insights for this investigation. This multitasking theory proposes that the limited capacity of working memory creates a bottleneck when resources are exhausted [7]. The bottleneck leads to a reduction in working memory resources for a primary task, when a secondary task of the same modality is introduced. As Wickens [6] suggests, when such homogeneous tasks are imposed, performance declines due to mental workload overload. For example, a visual display on a car stereo (a secondary task) may limit the resources available for the visual needs of the (primary) driving task. The theory suggests that in this instance, secondary tasks should be done via a different modality [6]. One potential modality in this case is the auditory modality.

Auditory displays for In-Car Infotainment Systems (ICIS) have been shown to increase a driver's visual attention on the primary driving task (driving), as compared to visual in-vehicle displays [8]. Results from a driving simulator and eye-tracking experiment showed sonification assistance, with respect to the ICIS, significantly reduced eye-movements towards the ICIS, thereby reducing participants' reaction times in the primary driving task. This finding corroborates the multiple resource theory model of multitasking by enhancing driving performance with reduced visual tasking [8]. Previous research has also resulted in similar conclusions for interfaces that used sound within the vehicle context, including increased visual attention on the driving task [9] and better driving performance [10].

While no audio-only FEDI is found in the literature or commercially, a FEDI including complimentary audio to a visual display has been studied [4]. That prototype multimodal display included a lower frequency tone (512 Hz) to indicate insufficient acceleration and a higher frequency tone (predominantly at 1770 Hz) to indicate excessive acceleration. Preferences revealed that participants tend to report displays with complementary audio as more effective at advising eco-driving behavior than visual-only displays [4]. In addition, behavioral measures showed that time spent looking at the road increased and drivers' pedal error (having the pedal outside the ideal range) decreased when using a system with added audio [4].

Unlike visual displays, auditory displays have the added capability to convey information to the driver, regardless of head or body position [11]. Auditory displays also allow for a wide range of information to be communicated to a driver through many dynamic acoustic parameters (in addition to speech sounds): frequency, timbre, range, register, and rhythm [11]. As Nees and Walker [12] suggest, an empirical investigation could determine the best mapping, scaling, and polarity of such sound features for a FEDI. Mapping, scaling, and polarity must be optimized to ensure that workload is not increased as a result of added auditory displays [13]. Further, driving performance and workload can be affected by annoyance [13]. When mapping acoustic parameters, it is important to consider trade-offs involving the effect of annoyance associated with some sounds [12].

2.1.1. HUDs

In-vehicle visual displays inherently demand more visual scanning time, thereby increasing cognitive load and distracting users from the driving task [14]. Empirical research has emphasized the importance of visual display placement in vehicles. The lower a display is positioned vertically (i.e., the farther below the windshield it is), the more severely driving performance is decreased, seen through increased reaction time and decreased target detection performance [15]. However, lane position can be maintained, even when attention is focused on in-vehicle displays, if the distance from the display to the outside line of sight is minimal [16]. This finding suggests that drivers can learn to manage dual-task load using peripheral vision, allowing them to maintain lane performance. Therefore, issues experienced using traditional visual displays may be overcome using head-up displays (HUD), which project information onto the vehicle windshield [17].

Simulator studies found that under both low- and high-difficulty driving conditions, drivers exhibit faster reaction times to task-related detection, such as speed limit sign changes, while attention is focused on the HUD [18]. Driving performance measures, such as variance of lateral acceleration, steering wheel turning (degrees), and acceleration are also improved while attending to the HUD as opposed to traditional head-down displays [18].

When designing HUDs, it is important to consider the trade-off between too much clutter and scanning-time cost. If too much information is displayed in the HUD, far-field vision becomes compromised and may cause attentional tunneling, which will decrease driving performance [19]. Therefore, it is recommended that the visual design reduce clutter by only including information that is pertinent to the task. As with all displays, ill-informed HUD designs could add to distraction, increased workload, and confusion [17]. Likewise, it is important to investigate the best mapping, scaling, and polarity of HUD designs to ensure the visual displays match a user's mental model of the system [20].

The amount of information communicated through visual displays has been studied in the context of FEDIs. A comparison of three different visual designs found that displays with greater information content were judged as more supportive for eco-driving behavior [4]. In that study a "foot-and-pedal" display showed current pedal error; a gauge display showed the rate of change of pedal error; and a dot display showed pedal error only. Results from behavior tests revealed speed maintenance with the gauge display was better than with the foot, whereas acceleration performance was better with the foot than with the gauge.

3. THE CURRENT RESEARCH

The primary goal of this study is to find patterns and preferences in the results that aid both auditory and HUD design decisions for future displays. The design guidance and prototypes this study yields for future research could result in advancements for both FE and driver safety. The present study suggests current visual heads-down eco-driving interfaces impose an additional visual demand to the already visually heavy task of driving. Although Young, Birrell, and Stanton [14] called for the development of eco-driving displays that decrease visual distraction, few examples of such research exist. There is a need to investigate the design of in-vehicle auditory and HUD displays that safely communicate how driver behavior affects fuel efficiency [2].

3.1. Types of Displays

3.1.1. Types of information displayed

There are two non-mutually exclusive categories of information that the displays in this study fall under: continuous and intermittent, and inform and instruct.

3.1.1.1 Continuous and Intermittent Displays

The optimal temporal structure of a display system is an important aspect of design. In a recent study of truck drivers' preference regarding visual FEDIs, both *continuous* and *intermittent* display prototypes were tested in a simulator [21]. Participants showed unique preferences for continuous and intermittent display types alike. A majority of participants noted that speed guidance (continuous display) was useful, easy to understand, and made controlling speed easier. A majority also found the performance feedback (intermittent display) positive, liking the feedback as an incentive to drive "eco-friendlier" [21]. In terms of an auditory display, a constant or continuous sound representing FE may generate annoyance [12]. Therefore, the current study assesses the best design parameters to be used in both a continuous and an intermittent display type for each modality (visual and auditory). For the purpose of this study, intermittent displays are designed to communicate the overall FE of a driving trip. In contrast, continuous displays are designed to give dynamic information about the current fuel economy.

3.1.1.2 Inform and Instruct Displays

In both the auditory and visual domain, this study proposes two primary display types for conveying information to the operator: *inform* and *instruct* displays.

Inform displays present information about a user's current behavior, increasing situation awareness to their current performance, allowing them to shape their behaviors to fit the task goals. In the current research domain, such a display might tell the driver if he or she is accelerating too fast or too slowly to meet the eco-driving goals.

On the other hand, *instruct* displays directly communicate how the user should change their behavior to accomplish the task goals. This means that users do not need to use this information to compare to the goal state but are instead told exactly what to do. So, for someone trying to increase their fuel economy an instruct display might tell a driver to accelerate more slowly.

Previous research has investigated a similar concept, finding that displays presenting more persuasive information (displays focused on convincing the user to change their behavior) were perceived as less useful and more difficult [22]. The instruct displays in this study (i.e., telling drivers how to behave), are analogous to those persuasive displays.

3.2. Experiment 1 (Auditory Matching) Overview

This participatory design study iteratively investigated sound parameters for the design of an auditory eco-driving interface. Sound-sorting methods provide an efficient way to categorize and evaluate sound design parameters, especially when there is a large number of stimuli [11, 23]. Participants matched sound parameters to eco-driving icons and descriptions. Participants completed a forced choice matching task with three sections (one for each display type): inform, instruct, and overall FE. Inform and instruct eco-driving concepts were investigated as continuous methods of display, whereas overall FE was an intermittent display. The frequency data collected through the number of times a sound was matched with a description of an eco-driving behavior established acoustic mappings, scaling, and polarity for each type of eco-driving display. In addition, a survey after each section of the forced choice task determined preferences for display types.

With the results of Experiment 1, the auditory design process took into account certain capabilities and limitations of sound characteristics inside the vehicle. A key limitation in vehicles is acoustic masking related to vehicle and road noise [12]. Experiment 1 results and special acoustic considerations facilitated the design of an intuitive and unambiguous auditory interface prototype to be used in future driving simulator research.

3.3. Experiment 2 (Visual Matching) Overview

This participatory design study investigated visual parameters for an iterative design of a HUD eco-driving interface. Card-sorting methods were also used as a way to categorize and evaluate visual designs. In the card-sorting, visual icons were matched to eco-driving words or concepts. There were three forced-choice matching tasks: instruct, inform, and overall FE. The frequency data collected through the number of times a sound was matched with a description of an eco-driving behavior established mappings, scaling, and polarity for each type of eco-driving display. Again, a survey following each task determined preferences for each type of display: instruct, inform, and overall FE.

In addition to the Experiment 2 data, the HUD design process took into account special visual design considerations, such as compromising far-field vision and attentional tunneling [17]. Experiment 2 results and special design considerations facilitated the design of an intuitive and usable HUD prototype for future driving research.

3.4. Research Questions and Hypotheses

Research Question 1: Which acoustic parameters are most useful and preferred for an auditory display of eco-driving concepts?

Research Question 2: Which visual parameters are most useful and preferred for a HUD of eco-driving concepts?

Hypothesis 1: Participants will display a higher preference for *inform* compared to *instruct* displays.

4. EXPERIMENT 1

4.1. Participants

Participants included 41 students (19 male) with an average age of 20.2 years (SD=1.8). Participants were required to be 18 or older to ensure they had some driving experience and were required to have normal or corrected to normal vision and hearing to control for abilities needed to perform the sound card-sorting task. They held a driver's license for an average of 3.7 years (SD=2.0).

4.2. Materials

The sound-sorting program used in the current study was written using HTML, jQuery, and Bootstrap's framework. Using Ableton Live music production software, sounds themselves were designed for acoustic parameters of interest. Surveys were constructed and executed online. Sounds were heard through SONY MDR-V150 Dynamic Stereo Headphones.

Three types of auditory displays were used in the study: earcons, auditory icons, and speech. *Earcons* are abstract sounds with no prior association (e.g., musical phrases), but the matching frequencies were expected to yield relationships between acoustic parameters and perceptions of constructs [11, 23]. *Auditory icons* are sounds that are ecologically associated (e.g., engine noises), and *speech* consisted of text-to-speech generations [12].

4.3. Procedure

Upon arrival participants signed consent forms and then sat at a computer, put on headphones, and began the sound-sorting program. The program started with a tutorial so that participants could become accustomed to the drag-and-drop procedure used to sort sounds.

There were three within-subject trials: one each for informational, instructional, and overall FE displays. The independent variables were the acoustic parameters presented on each slide: ADSR (attack, decay, sustain, release), audio effects, instruments (timbre), auditory icons, triads (musical chords), voices, scales (speech), register (musical octaves), range (distance between frequencies), and tempos. Sound parameters varied randomly within each trial. The first and second trials were randomly assigned. These trials investigated acoustic parameters for a continuous display. One trial asked participants to match sounds to driving *instruction* icon-word pairs, shown in Figure 2. Second, sounds were matched to icon-word pairs that *inform* a driver of current driving behavior status. A third trial investigated acoustic parameters for an intermittent display by having participants match sounds to a metric of overall FE.

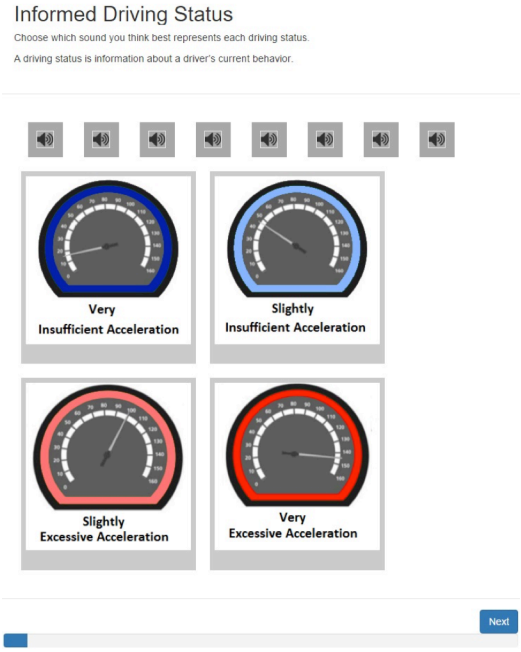


Figure 2: Inform display trial from Experiment 1.

Next, a survey asked participants which type of auditory displays (earcons, auditory icons, or speech) they believe to be most informative, least distracting, and would prefer. After both trials one and two were completed, there was an additional survey assessing user preference for an *inform* versus an *instruct* display. Following the third trial, a survey asked participants which type of overall FE display they believed to be most informative, least distracting, and would prefer: earcons, auditory icons, speech, earcons and speech, or auditory icons and speech.

4.4. Results

4.4.1. Sound card sorting results

Frequencies and percentages for sound-to-concept matches were calculated for the various acoustic parameters in each category. Due to the exploratory nature of the study, frequency counts were found to be the most useful type of data. Not all acoustic parameters resulted in high matching agreement between participants, but for some there were clear trends. As seen in Figure 3, there was high agreement among listeners that a musical phrase starting at note C0 and ascending to note C5 matched best to the concept of *accelerate a lot*, whereas a descending phrase from note C5 to C0 matched best to the concept of *decelerate a lot*. Similarly, C1 ascending to C4 and C2 to C3, matched best to *accelerate a little*, whereas C3 descending to C2 and C4 to C1 matched best to *decelerate a little*. The trend found in this particular set of sounds reveals auditory design guidance: an accelerate instruction is best matched to an increasing frequency; a decelerate instruction is best matched to a decreasing frequency; and range (distance between notes) can represent the magnitude of change in acceleration instructions. Using frequencies recorded in the three trials, an equivalent analysis was conducted for each acoustic parameter and auditory display studied: ADSR, audio effects, instruments, auditory icons, triads, voices, scales, register, range, and tempos. High matching agreement was determined and recorded (Table 1).

	Instruct	Inform	Overall Eco Preformance
ADSR	Accelerate	Insufficient Acceleration	Low Fuel Economy
	Slow Attack, High Sustain	Fast Attack, Low Sustain	High Sustain
	Decelerate	Excessive Acceleration	High Fuel Economy
	Fast Attack, Low Sustain	High Sustain	Fast Attack
Effects	Accelerate	Insufficient Acceleration	Low Fuel Economy
	High Freq. EQ	Reverb, Low Freq. EQ	Delay, Reverb
	Decelerate	Excessive Acceleration	High Fuel Economy
	Low Freq. EQ, Delay	Delay, Distort	High Freq. EQ
Instru-ments	Accelerate	Insufficient Acceleration	Low Fuel Economy
	String, Xylophone	Bass	Brass, Marimba
	Decelerate	Excessive Acceleration	High Fuel Economy
	Bass, Marimba	Brass	Xylophone
Icons	Accelerate	Insufficient Acceleration	Low Fuel Economy
	Engine Freq. Up	Engine Freq. Down	Short Guzzler
	Decelerate	Excessive Acceleration	High Fuel Economy
	Engine Freq. Down	Engine Freq. Up	Long Guzzler
Triads	Accelerate	Insufficient Acceleration	Low Fuel Economy
	Major Ascending	Aug. & Major Descending	Aug. & Dim. Descending
	Decelerate	Excessive Acceleration	High Fuel Economy
	Minor Descending	Major & Aug. Ascending	Major Ascending
Voices	Accelerate	Insufficient Acceleration	Low Fuel Economy
	Female	Female	Female
	Decelerate	Excessive Acceleration	High Fuel Economy
	Male & Female	Male & Female	Female
Scales	Accelerate	Insufficient Acceleration	Low Fuel Economy
	Major & Minor Ascending	Major Descending	Minor & Blues Descending
	Decelerate	Excessive Acceleration	High Fuel Economy
	Major & Minor Descending	Minor Ascending	Major & Minor Ascending
Register	Accelerate	Insufficient Acceleration	Low Fuel Economy
	C4 > C3 > C2	C2 > C3 > C4	C2 > C3 > C4
	Decelerate	Excessive Acceleration	High Fuel Economy
	C2 > C3 > C4	C4 > C3 > C2	C4 > C3 > C2
Range	Accelerate	Insufficient Acceleration	Low Fuel Economy
	C0-C5 Ascending	C5-C0 Descending	C5-C0 Descending
	Decelerate	Excessive Acceleration	High Fuel Economy
	C5-C0 Descending	C0-C5 Ascending	C0-C5 Ascending
Tempos	Accelerate	Insufficient Acceleration	Low Fuel Economy
	Fast, Increasing	Decreasing, Slow	Slow, Decreasing
	Decelerate	Excessive Acceleration	High Fuel Economy
	Decreasing, Slow	Increasing, Fast	Fast, Increasing

Table 1: Auditory parameters resulting in high matching agreement between participants.

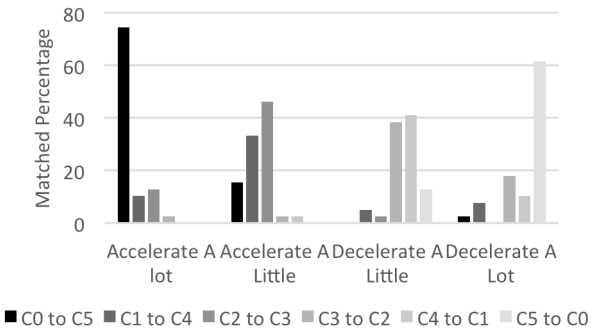


Figure 3: Match percentages of various ranges from Experiment 1 *instruct* trial.

4.4.2. Survey results

Response frequencies from the surveys following each trial show that auditory icons were generally reported as most distracting across all trials (see Table 2). The auditory icons were ecologically associated, meaning they resembled engine noises. This is compounded with acoustic masking and therefore, earcons and speech were considered best for an eco-driving auditory display. Participants were generally in agreement that speech was more informative than earcons in both the instruct and inform trials. However, there was no consensus as to which was preferred. In the overall FE trial, there was agreement that a display with *both* earcons and speech would be most informative and preferred.

Responses were higher for *instruct* displays over *inform* displays across all three measures. A chi square binomial probability test showed the *instruct* display had higher ease of understanding (p=0.001). This means that in auditory FEDIs instruct information was preferred.

Inform	**Earcon**	**Icon**	**Voice**
Most Informative	3	12	25
Least Distracting	18	8	14
Most Preferred	14	8	18
Instruct	**Earcon**	**Icon**	**Voice**
Most Informative	3	9	26
Least Distracting	16	6	16
Most Preferred	11	3	24

Overall Eco-Driving	**Earcon**	**Icon**	**Voice**	**Earcon & Voice**	**Icon & Voice**
Most Informative	4	2	8	19	6
Least Distracting	12	7	9	8	3
Most Preferred	8	1	8	15	7

Instruct vs Inform	**Instruct**	**Inform**
Ease of Use	30	10
Least Distracting	22	18
Preference	23	17

Table 2: Preference survey frequencies from Experiment 1.

5. EXPERIMENT 2

5.1. Participants

Participants were 46 students (24 males) with an average of 20.1 years old (SD=1.7). Participants had the same age, vision, and hearing requirements as in Experiment 1. Participants had held a driver's license for an average of 3.3 years (SD=1.9).

5.2. Procedure

Participants sat at a computer and began the icon-sorting program, starting with a tutorial to become familiar with the procedure. Participants then matched each icon to an eco-driving concept by dragging and dropping.

There were three within-subject trials. The independent variables were the types of icons presented on each slide: vertical continuous bars, vertical segmented bars, horizontal continuous bars, horizontal segmented bars, arches, up-down arrows, forward-back arrows, colors, leaves, trees, and shoe-on-pedal (Table 3). Images within each trial were randomly varied. The first and second trials were randomly assigned. These two trials investigated visual icons for a continuous display. One trial asked participants to match icons to driving *instructions*, as shown in Figure 4. In a second trial, icons were matched to words that *inform* a driver of current driving behavior status. The third trial investigated sound parameters for an intermittent display by asking participants to match sounds a metric of overall FE.

Following each of trials one and two, a survey asked participants which type of visual displays they believed to be most informative, least distracting, and would prefer as a user.

After both trials one and two were completed, there was an additional survey assessing user preference for an *inform* versus an *instruct* display. Subsequent to the third trial, a survey asked participants which type of overall FE display they believed to be most informative, least distracting, and would prefer as a user.

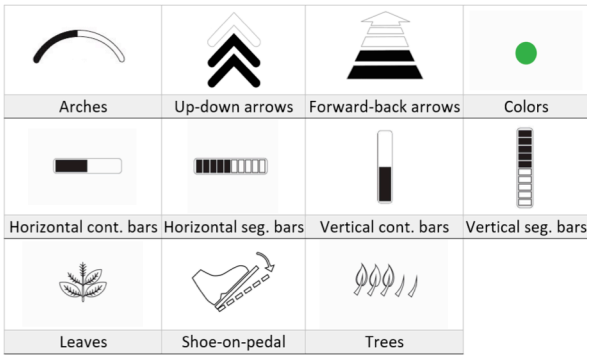


Table 3: An example of each type of image presented on the slides in Experiment 2.

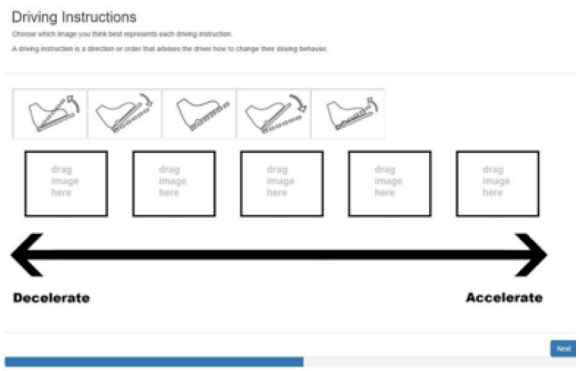


Figure 4: *Instruct* display trial from Experiment 2.

5.3. Results

5.3.1. Matching results

Frequencies and percentages for icons-to-concept matches were calculated for the various visual parameters in each category. It was not necessary to perform inferential statistical analyses because this study is exploratory in nature [23]. Most visual parameters did have high matching agreement between participants, indicating clear trends. As seen in Figure 5, there was high agreement among viewers that the color red matched best to the concept of *decelerate a lot*, the color orange matched best to the concept of *decelerate a little*, yellow matched best to *appropriate acceleration*, light green matched to *accelerate a little*, and dark green to *accelerate a lot*. Using frequencies recorded in all three trials, a similar analysis was conducted for each parameter and type of visual studied: vertical continuous bars, vertical segmented bars, horizontal continuous bars, horizontal segmented bars, arches, up-down arrows, forward-back arrows, colors, leaves, trees, and shoe-on-pedal. High matching agreement was recorded as seen in Table 4.

	Instruct	Inform	Overall Eco Performance
Vertical Cont. Bars		Insufficient Acceleration	Low Fuel Economy
		Down	Down
		Excessive Acceleration	High Fuel Economy
		Up	Up
Vertical Seg. Bars	Accelerate	Insufficient Acceleration	Low Fuel Economy
	Up	Down	Down
	Decelerate	Excessive Acceleration	High Fuel Economy
	Down	Up	Up
Horizontal Cont. Bars		Insufficient Acceleration	
		Left	
		Excessive Acceleration	
		Right	
Horizontal Seg. Bars		Insufficient Acceleration	
		Left	
		Excessive Acceleration	
		Right	
Arches		Insufficient Acceleration	
		Less fill	
		Excessive Acceleration	
		More fill	
Arrows (Forward/Back)	Accelerate		
	Forward		
	Decelerate		
	Back		
Arrows (Forward/Back)	Accelerate		
	Forward		
	Decelerate		
	Back		
Arrows (Up/Down)	Accelerate		
	Up		
	Decelerate		
	Down		
Colors	Accelerate	Insufficient Acceleration	Low Fuel Economy
	Green>Light Green>Yellow	Green>Light Green>Yellow	Red>Orange>Yellow
	Decelerate	Excessive Acceleration	High Fuel Economy
	Red>Orange>Yellow	Red>Orange>Yellow	Green>Light Green>Yellow
Leaves			Low Fuel Economy
			Less leaves
			High Fuel Economy
			More leaves
Trees			Low Fuel Economy
			Less trees
			High Fuel Economy
			More trees
Shoes/Pedals	Accelerate		
	Low angle		
	Decelerate		
	High angle		

Table 4: Visual parameters resulting in high matching agreement between participants.

5.3.2. Survey results

As shown in Table 5, in the overall FE trial, the leaf icons and horizontal segmented bars were seen as least informative, most distracting, and least preferred. In the inform trial, the arch icons were reported as most informative, least distracting, and most preferred. The foot-to-pedal icons were most distracting while the up-down arrow icons were preferred in the instruct trial. In all other measures there were no clear participant preferences.

Frequencies were higher for instruct displays over inform displays across all three trials. Chi square binomial probability tests of responses showed that the inform icons had significantly greater ease of understanding (p=0.024), were the least distracting (p<0.001), and most preferred (p<0.001).

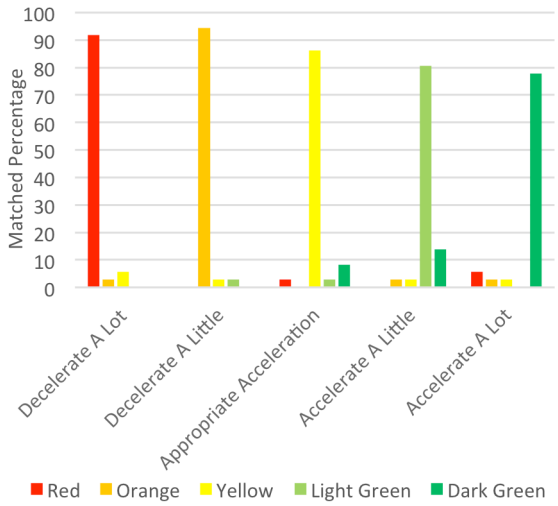


Figure 5: Match percentages of various colors from Experiment 2 instruct trial.

Inform	Arches	Vertical seg. bars	Horizontal seg. bars	Vertical cont. bars	Horizontal cont. bars
Most Informative	25	6	10	1	2
Least Distracting	21	8	6	4	5
Most Preferred	28	6	5	1	4
Instruct	**Arrows (up/down)**	**Arrows (left/right)**	**Vertical seg. bars**	**Foot & Pedal**	
Most Informative	13	10	8	13	
Least Distracting	15	9	15	5	
Most Preferred	17	10	10	7	
Overall Eco-Driving	**Tree**	**Leaves**	**Vertical seg. bars**	**Horizontal seg. bars**	
Most Informative	16	9	16	4	
Least Distracting	12	5	19	9	
Most Preferred	15	8	16	6	
Instruct vs Inform	**Instruct**		**Inform**		
Ease of Use	15		29		
Least Distracting	11		33		
Preference	9		35		

Table 5: Preference survey frequencies from Experiment 2.

6. GENERAL DISCUSSION

The results of Experiments 1 and 2 comparing *instruct* and *inform* display types expose a potential hurdle for design. Acoustic parameters that matched to *instruct* icon-word pairs generally revealed an opposite polarity to those matched to *inform* icon-word pairs (Table 1). For example, in the instruct trial, participants matched ascending pitches to accelerate, and descending pitches to decelerate. In the *inform* trial, participants matched ascending pitches to excessive acceleration and descending pitches to insufficient acceleration. Here the polarities are opposite because *instructing* a driver to accelerate and *informing* a driver of insufficient acceleration is delivering information about the same driving phenomenon. However, these two analogous concepts were mapped to opposite acoustic parameters: ascending and descending frequencies, respectively. Decelerate and excessive acceleration (similar concepts) were also mapped to opposite acoustic parameters: descending and ascending, respectively. This may actually make it simpler to develop auditory interfaces for this task, as the designers do not have to worry about issues of cross coding of the displays.

The same pattern can be seen in the Experiment 2 icon-to-concept matching results (Table 4). For example, a vertical bar in the up position matched best to accelerate, but was also matched to excessive acceleration. Similarly, a vertical bar in the down position matched best to decelerate, but was also matched to insufficient acceleration. The opposite polarities between inform and *instruct* displays could be problematic for a design, which draws from both. This issue must be considered in future multimodal displays.

Experiment 1 surveys found *instruct* auditory displays preferable, while Experiment 2 surveys found *inform* visual displays preferable (*Hypothesis 1*). However, a multimodal display using *instruct* sounds and *inform* visuals would present incongruent information with opposite polarities. This could lead to user confusion and distraction in the driving environment. These findings should also be used to advise future multimodal design.

6.1. Human-centric and System-centric Displays

Human-centric and system-centric displays are two approaches to showing information to drivers, whether it is information based on human parameters (human-centric), or system parameters (system-centric). In the current study both the instruct and inform displays were human-centric because they directly told the driver how to engage or change driving behavior or told them about their eco-driving behaviors, both of which were focused on the human. However, the overall FE information was a system-centric display as it displayed a metric describing the fuel economy within the system to the driver. These factors should be considered further in future work.

6.2. Major findings and future research

The purpose of this study was to find patterns and preferences that contribute to both an auditory and HUD design. Major findings include the high auditory and visual matching frequencies (Tables 1 & 4; *Research Question 1 & 2*). These trends, along with survey results indicating preferences, directed our design of intuitive, usable, and unambiguous auditory and HUD FEDIs. Future research will evaluate how research-driven designs compare to commercially used visual dashboard displays. This next-step research will investigate these prototypes in a driving simulator study. The primary measures will include eco-driving behavior, eye behavior, subjective workload, and driving performance. Future research should also consider other age groups to ensure that matching of the displays does carry over age groups, or to determine what differences are seen between age groups. Results from these works could serve as a verification of the design guidance in this study and will help determine what types of displays effectively and safely communicate fuel efficiency to shape driver behavior.

7. ACKNOWLEDGMENTS

We would like to thank Brianna Tomlinson, Abhishek Sen, and Yiwei Hao for their help with the research. Portions of the work were supported by a National Science Foundation Graduate Research Fellowship (DGE-1148903) as well as additional grant funding from the NSF and from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR).

8. REFERENCES

1. Staubach, M., Schebitz, N., Koster, F., & Kuck, D. (2014). Evaluation of an eco-driving support system. *Transportation Research Part F: Traffic Psychology and Behavior*.
2. Nees, M. A., Gable, T., Jeon, M., & Walker, B. N. (2014). Prototype auditory displays for a fuel efficiency driver interface. In *20th International Conference on Auditory Display*.
3. Gonder, J., Earleywine M., & Sparks, W. (2011). Final report on the fuel saving effectiveness of various driver feedback approaches. *National Renewable Energy Laboratory*.
4. Jamson, A., Hibberd, D., & Merat, N. (2015). Interface design considerations for an in-vehicle eco-driving assistance system. *Transportation Research Part C: Emerging Technologies.*
5. Waard, D. (1996). The measurement of drivers' mental workload. *Traffic Research Center: Groningen University.*
6. Wickens, C. (2008). Multiple resources and mental workload. *Human Factors,* 6, 449-55.
7. Ralph, J., Gray, W. D., & Schoelles, M. J. (2013) Cognitive Workload and the Motor Component of Visual Attention.
8. Tardieu, J., Misdariis, N., Langlois, S., Gaillard, P., & Lemercier, C. (2015). Sonification of in-vehicle interface reduces gaze movements under dual-task condition. *Applied ergonomics*, *50*, 41-49.
9. Gable, T. M., Walker, B. N., Moses, H. R., & Chitloor, R. D. (2013, October). Advanced auditory cues on mobile phones help keep drivers' eyes on the road. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 66-73). ACM.
10. Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu navigation with in-vehicle technologies: Auditory menu cues improve dual task performance, preference, and workload. *International Journal of Human-Computer Interaction*, *31*(1), 1-16.

11. Lewis, B., & Baldwin, C. (2015). Comparison of traditional psychophysical and sorting methods for in-vehicle display design. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting*.
12. Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of human factors and ergonomics*, *7*(1), 58-99.
13. Wiese, E., & Lee, J. D. (2001, October). Effects of multiple auditory alerts for in-vehicle information systems on driver attitudes and performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 45, No. 23, pp. 1632-1636). SAGE Publications.
14. Young, M. S., Birrell, S. A., & Stanton, N. A. (2011). Safe driving in a green world: A review of driver performance benchmarks and technologies to support 'smart' driving. *Applied ergonomics*, 42(4), 533-539.
15. Burns, P. C., Andersson, H., & Ekfjorden, A. (2001). Placing visual displays in vehicles: where should they go? Internal Conference on Traffic and Transport. Psychology-ICTTP 4-7 September 2000.
16. Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *38*(3), 442-451.
17. Wittmann, M., Kiss, M., Gugg, P., Steffen, A., Fink, M., Pöppel, E., & Kamiya, H. (2006). Effects of display position of a visual in-vehicle task on simulated driving. *Applied Ergonomics*, *37*(2), 187-199.
18. Liu, Y. C. (2003). Effects of using head-up display in automobile context on attention demand and driving performance. *Displays*, *24*(4), 157-165.
19. Guastello, S. J. (2013). Human factors engineering and ergonomics: A systems approach. CRC Press.
20. Sanders, M. S. & McCormick, E. J. (1993). *Human Factors in Engineering and Design* (7th ed.). New York: McGraw-Hill.
21. Fors, C., Kircher, K., & Ahlström, C. (2015). Interface design of eco-driving support systems– Truck drivers' preferences and behavioural compliance. *Transportation Research Part C: Emerging Technologies*, *58*, 706-720.
22. Meschtscherjakov, A., Wilfinger, D., Scherndl, T., & Tscheligi, M. (2009). Acceptance of future persuasive in-car interfaces towards a more economic driving behaviour. *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*.
23. Tomlinson B., Schuett, J. H., Shortridge, W. H., & Walker, B. N. (2016). "Talkin About the Weather": Incorporating TalkBack Functionality and Sonifications for Accessible App Design. In *Mobile human-computer interaction-MobileHCI 2016*

USE OF SONIFICATION OF RADAR DATA FOR NOISE CONTROL

Kees van den Doel

Adrok Ltd.
49-1 West Bowling Green Street
Edinburgh, EH6 5NX, Scotland
kdoel@adrokgroup.com

Michael Robinson

Adrok Ltd.
49-1 West Bowling Green Street
Edinburgh, EH6 5NX, Scotland
mrobinson@adrokgroup.com

ABSTRACT

Deep sounding radar surveys for geophysical exploration requires the detection of faint reflections from deep subsurface structures. Signal to noise enhancement through extensive data stacking is effective provided the data noise is incoherent and time-invariant. We describe the use of sonification of radar data for quality control of peripheral equipment, specifically to detect unwanted noise with a temporal pattern. The sonification process consists of filtering and time-scaling radio frequency data and interpreting the result as audio, a process usually referred to as auralization. A small user study was performed to quantify variations in individual performance in detecting these.

1. INTRODUCTION

Applications of ground penetrating radar [1] are currently mostly limited to shallow depths of a few tens of meters, because of the strong attenuation of radio waves in most subsurface materials at a typical frequency range of $15 - 1000$MHz. Losses are caused by conductivity and polarization effects due to moisture content or inherent material properties. Deeper penetration has been achieved with much lower frequencies ($1 - 5$Mhz) using very large antenna's in resistive environments such as Martian rock, ice, and permafrost [2, 3, 4].

https://doi.org/10.21785/icad2017.026

Adrok has developed a radar based imaging technology operating in a similar frequency range which has been available to the market for over five years with the express purpose of extending the depth range of conventional GPR surveys, in addition to introducing other novel methods such as spectroscopy [5, 6, 7]. Applications are mainly in geophysical exploration.

Detecting reflections from depths up to the kilometer range requires sensitive digital data recording peripherals attached to the antenna system, and extensive stacking ($200,000$ measurements are routinely performed) to increase the signal to noise ratio. Often the surveys are performed in remote hostile environments, see for example Figure 1, and the collected data is later analysed and interpreted. To make sure our data is as clean as possible we have a quality control protocol to check the integrity of the data, checking for possible equipment malfunctions on-site.

Recently we have started deploying data auralization methods to quickly check for possible problems, which usually appear as spurious noise from peripheral equipment mixing in with the radar data. Though we do not use any sophisticated auralization methods, we hope the ICAD community will be interested in this "real-world" application of auralization methods.

2. EQUIPMENT PERFORMANCE MONITORING

The weak reflections from subsurface reflectors require the best signal to noise ratio achievable, as the depth limit of the imaging is restricted by the signal to noise ratio. As such we use extensive data stacking, which means averaging repeated measurements, usually several hundred thousand. Similar stacking is used

Figure 1: A scene from field work in the arctic.

in seismic, but because the propagation speed of electromagnetic waves is many orders of magnitude higher than the speed of sound, a single measurement takes only tens of microseconds, which allows such large repeats.

The principle of stacking is that coherent signals (i.e., the same in every repeat) add up linearly (proportional to n, with n the number of measurements), but incoherent noise adds up to $\sqrt{n}$. Therefore the most "dangerous" equipment noise is coherent noise that is synchronized with the measurements, as it will not be relatively reduced through stacking.

During a recent survey in the Canadian arctic an anomaly was detected in the measured data, which was traced to a periodic coherent noise generated by a malfunction in a peripheral electronics component. Our standard visual tools were able to diagnose the problem in one data set, but on closer examination we found a similar problem but much weaker in some other data sets, but it was too weak to show up in standard visual plots like waveforms and spectrograms. However using data auralization the problem was clearly identified.

3. DATA AURALIZATION

Data is acquired in the form of digitized waveforms at a typical sampling rate of 5GHz with total duration of 20μs. Weak reflections from subsurface structures are detected after stacking using various signal processing methods. In most cases they are too weak to be detected using auralization of the waveform. Background and equipment noise is measured by taking "passive" measurements, with just the listening antenna active. If everything is working properly the result after stacking should be colored noise without any temporal structure. The passive stack is sonified in straightforward fashion by interpreting it as an acoustic wave in a suitable audio range, typically a reduction to a sampling rate of 25KHz. We then listen to the resulting audio to check the audio is time-invariant. Any external (i.e., not resulting from the antenna peripherals) disturbances if present will be incoherent across the recordings in the stack and thus will not result in localized temporal disturbances. Hence if we can hear any structure in the audio, there is a potential problem and further investigation and troubleshooting is initiated.

In Figure 2 we show an example of a signal with a periodic disturbance strong enough to be identified visually through a spectrogram, though not visible in the waveform depicted in Figure 3.

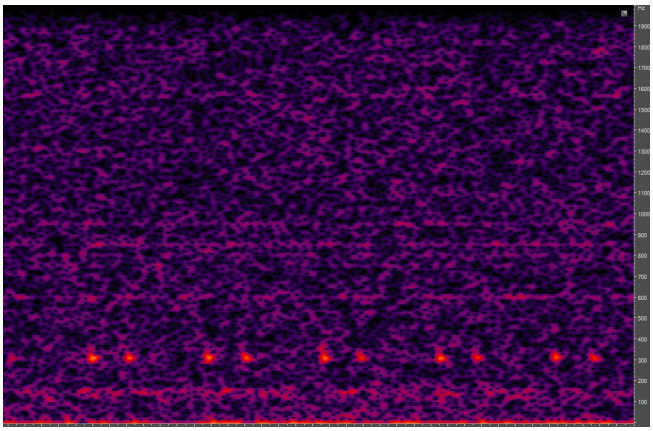


Figure 2: Spectrogram of a passive recording, scaled into the audio range. The periodic disturbance is clearly visible as well as audible. Audio file fig2.wav in supplementary data.

A more interesting example from the point of view of auralization is depicted in Figure 4. In this case the data looks fine visually, but auralization does reveal a problem similar as in the data corresponding to Figure 2 as it "sounds similar".

4. USER STUDY

It was noticed that some people found it difficult to diagnose subtle effects in the auralizations, even when it

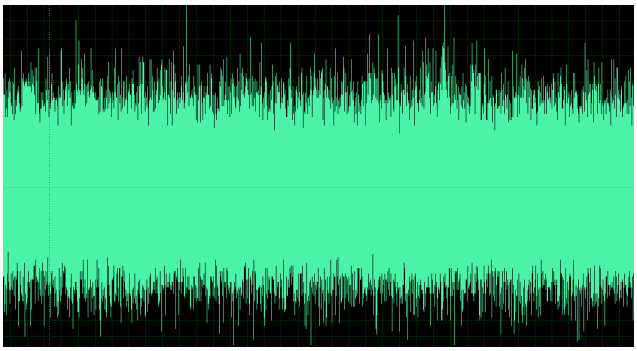


Figure 3: Waveform corresponding to Figure 2.

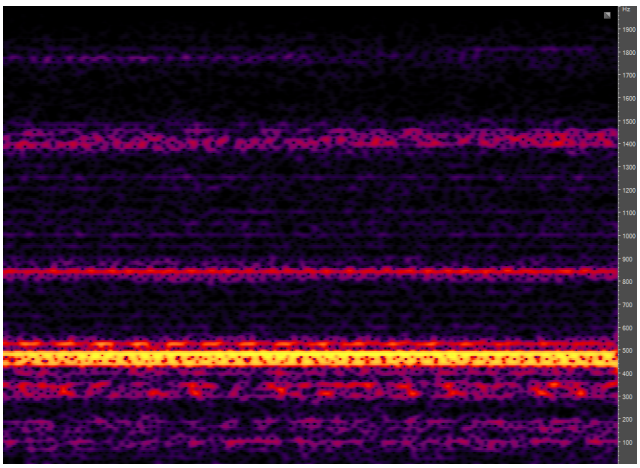


Figure 4: Spectrogram of passive recording, scaled into the audio range. The noise bands are known incoherent external sources and are not indicative of any problems. No periodic disturbance as in Figure 2 is seen but is audible when listening to the sound. Audio file fig4.wav in supplementary data.

was obvious to others. We performed a small (15 subjects) user study to determine how performance varies over individuals. Subjects were instructed to listen to 15 auralizations, 7 of which had an anomaly present. Three examples were given of auralizations with no, faint, and clear periodic disturbances (anomaly). All examples were taken from actual field work. Results are summarized in Figure 5. 8 out of 15 subjects did not miss any anomalies, and 4 subjects had no errors. As expected the faintest anomaly performed worst with only 73% of the subjects detecting it and the clearest anomaly was detected by all subjects.

Clearly there are significant variations between people, all with normal hearing. Probably the detection of anomalies does not depend directly on the frequency selectivity of the ear, but more on the ability to perform auditory scene analysis [8]. It remains to be investigated if training can help improving the ability to detect anomalies.

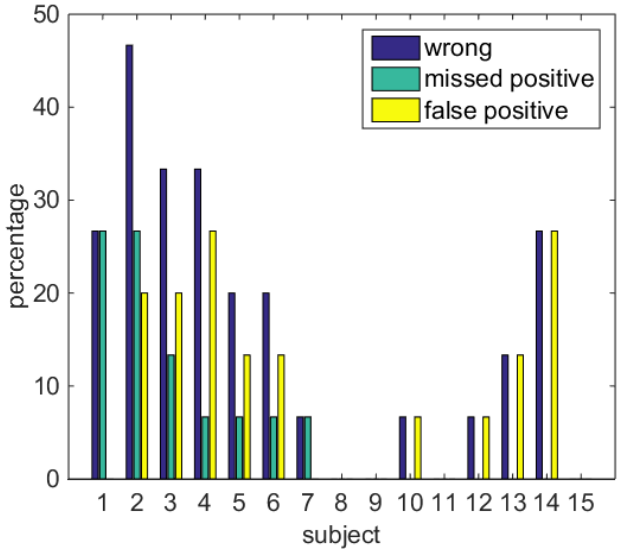


Figure 5: Score per subject indicating wrong answers, missed anomalies, and false positives. Note subjects 8, 8, 11, and 15 had a perfect score.

5. CONCLUSIONS

The detection of weak subsurface radar reflections is depth limited by the signal to noise ratio. The fast acquisition rate of a radar "shot" allows extensive stacking for denoising, provided the background noise is time invariant. In the often harsh conditions encountered during field surveys occasional equipment malfunctions are unavoidable and we have found auralization techniques to be an excellent tool to quickly and reliably identify non time invariant disturbances in our data sets.

We performed a small user study and found that about half the subjects are able to use the auralizations successfully (meaning no missed anomalies and only a few false positives) for anomaly detection. A false positive is not a problem, as further investigation would reveal if the problem is real or not, but a missed anomaly would cause an equipment problem to go unnoticed. Whether training improves the performance is an open question.

We plan to further refine these techniques and collect reference audio signals of various equipment so

any change in quality of the sound can be used as a warning sign of a potential problem during field work.

6. SUPPLEMENTARY DATA

For audio examples referenced in the figures, see: https://dl.dropboxusercontent.com/u/7482624/icad2017.zip

7. REFERENCES

[1] H. M. Jol, *Ground Penetrating Radar Theory and Applications*. Amsterdam: Elsevier, 2009.

[2] J. J. Berthelier, S. Bonaime, V. Ciarletti, R. Clairquin, F. Dolon, A. L. Gall, D. Nevejans, R. Ney, and A. Reineix, "Initial results of the Netlander imaging ground-penetrating radar operated on the Antarctic Ice Shelf," *GEOPHYSICAL RESEARCH LETTERS*, vol. 32, no. L22305, pp. L22 305, doi:10.1029/2005GL024 203, 2005.

[3] M. C. Angelopoulos, W. H. Pollard, and N. J. Couture, "The application of CCR and GPR to characterize ground ice conditions at Parsons Lake, Northwest Territories," *Cold Regions Science and Technology*, vol. 85, pp. 22–33, 2013.

[4] R. Jordan, G. Picardi, J. Plaut, K. Wheeler, D. Kirchner, A. Safaeinili, W. Johnson, R. Seu, D. Calabrese, E. Zampolini, A. Cicchetti, R. Huff, D. Gurnett, A. Ivanov, W. Kofman, R. Orosei, T. Thompson, P. Edenhofer, and O. Bombaci, "The Mars express MARSIS sounderinstrument," *Planetary and Space Science*, vol. 57, pp. 1975–1986, 2009.

[5] K. v. d. Doel, J. Jansen, M. Robinson, G. C. Stove, and G. D. C. Stove, "Ground penetrating abilities of broadband pulsed radar in the 1-70MHz range," in *SEG Technical Program Expanded Abstracts 2014, Denver*, 2014, pp. 1770–1774.

[6] G. Stove and K. van den Doel, "Large depth exploration using pulsed radar," in *ASEG-PESA Technical Program Expanded Abstracts 2015, Perth*, 2015, pp. 1–4.

[7] K. van den Doel, "Modeling and Simulation of a Deeply Penetrating Low Frequency Subsurface Radar System," in *78th EAGE Conference and Exhibition 2016, Vienna*, 2016, pp. doi: 10.3997/2214–4609.201 601 033.

[8] A. S. Bregman, *Auditory Scene Analysis*. Cambridge: The MIT Press, 1990.

Paper Session 4
Ecology

ILLUSTRATING TRENDS IN NITROGEN OXIDES ACROSS THE UNITED STATES USING SONIFICATION

Joshua L. Laughner

Department of Chemistry
University of California, Berkeley
Berkeley, CA 94720 USA
jlaughner@berkeley.edu

Elliot Kermit Canfield-Dafilou

Center for Computer Research
in Music and Acoustics
Stanford University, Stanford, CA 94305 USA
kermit@ccrma.stanford.edu

ABSTRACT

Leveraging the human auditory system, sonification can be used as an educational tool for non-experts to engage with data in a different mode than visualization. Without oversimplifying the data, this project presents a sonification tool for exploring NO_2 and O_3 data from the BErkeley High Resolution (BEHR) tropospheric NO_2 and OMO3PR ozone profile datasets. By allowing the listener control over the data-to-sound mapping and synthesis parameters, one can experience and learn about the interplay between NO_2 tVCDs and O_3 concentrations. Furthermore, interannual trends can be perceived across different types of locations.

1. INTRODUCTION

1.1. Nitrogen oxides play a key role in controlling air quality.

Nitrogen oxide (NO) and nitrogen dioxide (NO_2), collectively known as NO_x, play an important role in air quality. Photolysis of NO_2 produces ozone (O_3), and the reaction of NO with oxidized volatile organic compounds (VOCs) can lead to the formation of fine particulate matter.

Both ozone and particulate matter concentrations in the atmosphere are regulated by the Environmental Protection Agency (EPA) because of the negative health effects associated with exposure to them. Elevated concentrations of both are known to cause respiratory distress, especially in children [1]. Elevated ozone also damages crops, leading to significant economic losses as well as reducing food yields [2].

The role NO_x plays in the production of O_3 is complex, as the production efficiency of O_3 depends nonlinearly on both the NO_x concentration and the concentrations and identities of VOCs present in the atmosphere. At both high and low concentrations of NO_x, O_3 production due to NO_x cycling is suppressed, though for different reasons. At intermediate NO_x concentrations, ozone production peaks [3]. Therefore, cities attempting to improve their air quality by reducing NO_x concentrations may see an increase in ozone initially, and a decrease only once NO_x concentrations have fallen below a critical point. The value of that critical point depends on the mixture of VOCs present in the atmosphere.

NO_x is emitted through a number of anthropogenic and natural processes. Anthropogenic sources are typically those involving combustion, as the high temperatures break the N_2 and O_2 molecules in the atmosphere, allowing them to recombine as NO or NO_2. Examples of such sources are vehicles, power plants, ships, and aircraft. Natural sources include high temperature sources, such as biomass burning or lightning, as well as other sources such as soil bacteria [4].

1.2. Space-based measurement of NO_2 and O_3 offer broad geographic and temporal coverage

Space-based measurements of NO_2 tropospheric column density began over two decades ago with the launch of the Global Ozone Monitoring Experiment (GOME) instrument on board the ERS-2 satellite in 1996 [5]. Only NO_2, rather than total NO_x is measured due to its spectroscopic properties. Since then, several additional instruments have been launched, including the SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY (SCIAMACHY) [6], Ozone Monitoring Instrument (OMI) [7], and GOME-2 [8]. All these instruments are carried on board polar orbiting satellites which allows them to observe the entire globe in 1–6 days, depending on the instrument and operational mode.

Space-based observations of NO_2 offer a level of combined spatial and temporal coverage not possible with ground- or aircraft- based instruments. This offers several notable advantages, such as the ability to observe an entire urban and suburban area, to compare multiple urban areas across the globe using the same instrument, and the ability to monitor episodic events (biomass burning, lightning) difficult to track with other types of instruments. Multiple papers have made use of these properties to investigate both anthropogenic [9–18] and natural NO_x emissions [19–24].

The result of these measurements is a "tropospheric vertical column density" (tVCD), usually in units of molecules/cm^2. This is the total number of molecules of NO_2 over one square centimeter of the Earth's surface between the surface and the top of the troposphere (typically $\sim$ 12 km). Rural areas considered "clean" typically have tVCDs of $\leq 1 \times 10^{15}$ molec. cm^{-2}. Highly polluted areas such as Los Angeles, CA, USA or Beijing, China have tVCDs in excess of 1×10^{16} molec. cm^{-2}.

Measurements of O_3 from space can be done similarly to measurements of NO_2 using ultraviolet-visible spectroscopy and are almost always measured by the same satellites. However, measurements of *tropospheric* O_3 are complicated by the high concentration of O_3 in the stratosphere. Whereas the tropospheric and stratospheric components of the total NO_2 vertical column density are similar orders of magnitude, the tropospheric compo-

https://doi.org/10.21785/icad2017.016

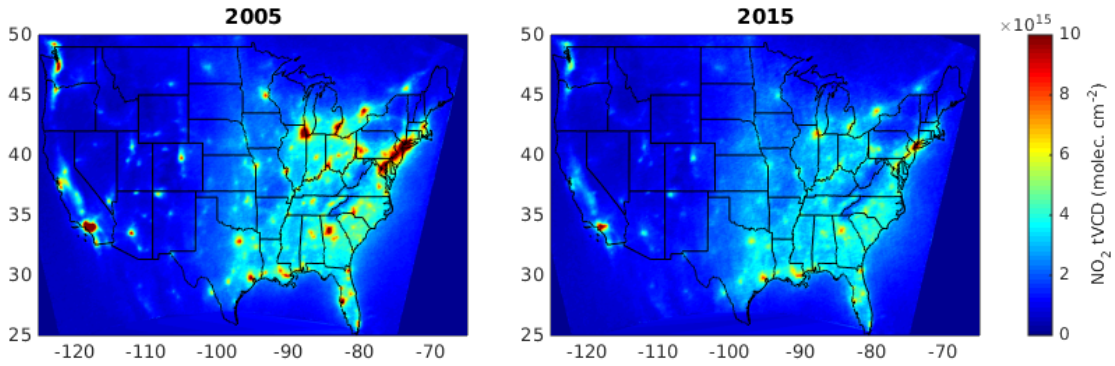


Figure 1: NO_2 summertime (Apr–Sept) tVCDs from the BEHR product for 2005 and 2015. A clear decrease across the US can be seen.

nent of the O_3 total VCD is minor compared to the stratospheric component [4]. Alternatively, the retrieval of tropospheric O_3 may be done using infrared spectroscopy [25].

1.3. NO_x has decreased in the US over the past decade.

In the US, the Environmental Protection Agency (EPA) has regulated measures to decrease the emissions of NO_x in order to reduce tropospheric ozone concentrations [26]. Regulations targeted both vehicular emissions [27] and power plant emissions [28]. Satellite observations of NO_2 [16, 18, 29] can clearly see the decrease in NO_2 throughout the US for the time period 2004 onwards (Fig. 1).

1.4. Sonifcation is an ideal educational tool to communicate the complexity of NO_x/O_3 chemistry.

NO_x and O_3 have interesting temporal patterns on both the interannual and seasonal time scales. As discussed in §1.3, NO_x has, in general, decreased across the US in the past decade. NO_x tVCDs also follow a seasonal cycle, primarily due to temperature dependent shifts in the chemistry. This leads to a sinusoidal pattern superimposed on top of the interannual decrease. The chemistry described in §1.1 means that O_3 concentrations will be related to NO_x tVCDs, but the exact dependence will vary from location to location.

This is shown in Fig. 2 for two cities, one power plant, and one rural area. Both seasonal and interannual trends are clear. The cities and power plants exhibit maximum NO_2 values in the winter while the rural location does so in the summer. Overall, NO_2 tVCDs decrease over the cities and power plant while remaining fairly constant over Yellowstone. O_3 appears to increase over the interannual time scale for the cities and power plant, but not over Yellowstone.

These characteristics make sonification an ideal way to describe the NO_x/O_3 relationship throughout the US. The temporal dependence of the data lends itself naturally to depiction in a time-dependent medium such as sound. The geographically diverse natural of the dataset can be well represented by the placement of sound in the panning field. By simultaneously representing the NO_2 tVCDs and O_3 concentrations at multiple cities, power plants, and rural areas across the US, we provide an intuitive interface for the public to learn about how reductions in NO_x concentrations affect O_3 under different conditions.

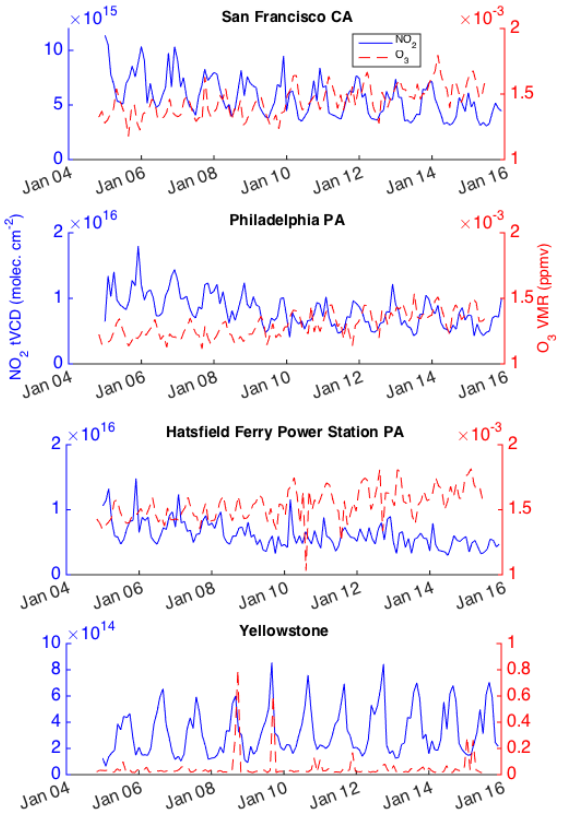


Figure 2: NO_2 tVCD and O_3 surface concentration trends at two cities (San Francisco, CA and Philadelphia, PA), one coal-fired power plant (Hatsfield, in Masontown, PA) and Yellowstone National Park (WY).

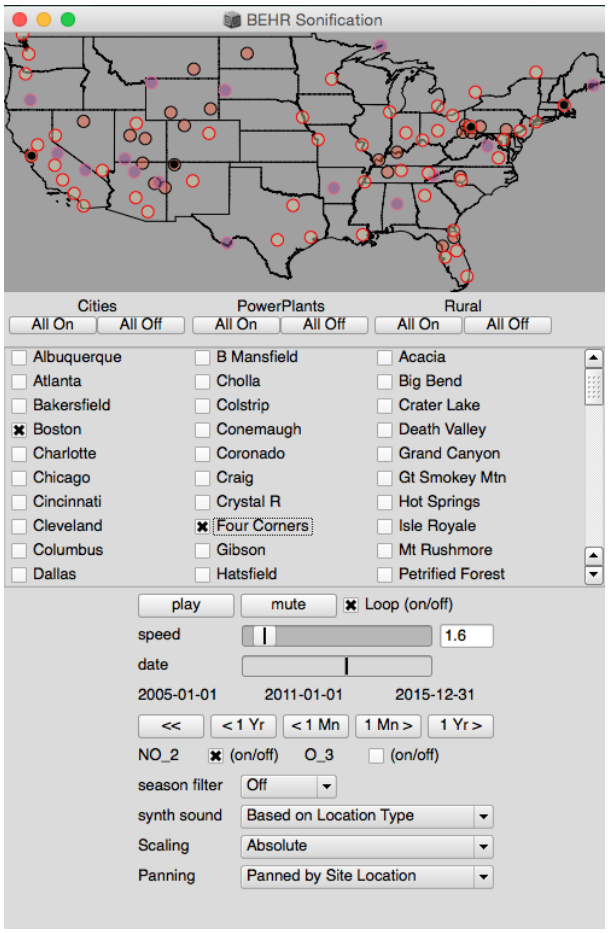


Figure 3: BEHR Sonification GUI.

2. APPROACH TO SONIFICATION

As an educational tool, sonification allows the user to engage with data through completely different modes than visualization. The streaming capabilities of the human auditory system means that it is good at processing multiple, synchronous data series. As §1.1 describes, the interactions of atmospheric chemicals, meteorological conditions, and ground activity are extremely complicated. Nevertheless, at various time scales, the trends in NO_2 and O_3 can be intuitively understood through the auditory experience. Unlike data-music, this sonification project has distinct educational goals and is designed for non-experts in particular. While we strive to reduce the complexity of the data, we want the sonification model to convey useful information. Furthermore, we want an interface that gives the user flexibility to determine how the data should be presented. This fulfills the goal to let the user explore the data in a way that conveys pertinent information.

2.1. NO_2 satellite dataset

We make use of v2-1C of the BErkeley High Resolution (BEHR) Ozone Monitoring Instrument (OMI) NO_2 gridded product, which is publicly available at `http://behr.cchem.berkeley.edu/DownloadBEHRData.aspx`. The BEHR dataset is chosen because it uses high-resolution *a priori* NO_2 profiles that better resolve the urban/rural NO_2 gradient than the NASA Standard Product or the KNMI DOMINO product. The OMI is carried on board the NASA Aura satellite, launched in 2004, and is a nadir-viewing, UV-visible spectrometer with an overpass time of 13:30–14:00 local standard time [7].

We use the cities and power plants identified in Russell et al. 2012 [16] as the sites for urban and power plant trends. We choose 15 additional sites in rural areas, focusing mostly on national parks, to demonstrate NO_2 variability in areas substantially less influenced by anthropogenic emissions. The radii for these sites are set to 40 km; this choice is arbitrary, as there is no clear plume to encompass, but is similar to the average radius used for the cities. Monthly average NO_2 tropospheric vertical column densities (tVCDs) are used to generate the trends. First, the gridded product is restricted to data meeting the following criteria:

- Cloud fraction ≤ 0.2
- The XTrackQualityFlags value must be 0 for all pixels that contribute to this grid cell
- The vcdQualityFlags must be an even integer (least significant bit is 0)
- Only rows 1–58 (0 based indexing) are used due to an issue in v2-1C of the BEHR product that causes the edge rows to be too large.

The gridded data is temporally averaged, weighted by the inverse of the pixel areas that contribute to the grid cells. This gives more weight to smaller, more representative pixels. For each monthly average, the grid cells whose centers are within the radius of the site longitude and latitude given in [16] are then themselves averaged to give a single value for each site for each month.

2.2. O_3 satellite dataset

We use the OMO3PR ozone profile dataset to obtain tropospheric O_3 available from `https://disc.gsfc.nasa.gov/Aura/data-holdings/OMI/omo3pr_v003.shtml`. This retrieval uses optimal estimation to fit an O_3 profile to observed absorbance in two wavelengths. An *a priori* O_3 profile is used as the basis for the profile shape. We choose this product because it is also derived from the Ozone Monitoring Instrument, as is our NO_2 product, and so has similar spatial and temporal coverage. Kroon et al. compared this product against multiple other satellite O_3 products and *in situ* sonde measurements and found a +30% bias in the midlatitudes [30]. While this alone should not interfere significantly in the use of this product for the trend sonification in this work as the bias is systematic, uncertainty in the values will be high due to the challenge of separating tropospheric and stratospheric O_3.

A gridded version of this product is not available; therefore we obtain monthly averages differently than with the NO_2 product. Similarly to NO_2, pixels with centers within the radii defined for the geographic sites are identified as contributing to the trend for that site; unlike the NO_2 product, these are the native satellite pixels, rather than regridded data. All valid pixels for a site for a month are binned in this step. Pixels are considered invalid if:

- Cloud fraction for either UV channel is > 0.2
- Aerosol optical thickness is $> 10^{-5}$. Elevated aerosol layers lead to erroneously large tropospheric O_3 concentrations [31].

- The ReflectanceCostFunction field is > 30. This indicates erroneous radiance due to the OMI Row Anomaly [32].

For each binned profile, two quantities are calculated. First, the bottom partial column (in Dobson Units) is converted into volume mixing ratio in units of part-per-million by volume (ppmv) using the formula provided in the OMO3PR Readme [31]. Because this formula relies on the edges of the pressure bin and the lower edge is the surface, we calculate the surface pressure for each profile by interpolating the Global Land One-km Base Elevation (GLOBE) project elevation data [33] to the pixel latitude and longitude, then convert from altitude to pressure using Eq. (1):

$$p = p_0 e^{-z/H} , \tag{1}$$

where p_0 is the sea level pressure of 1013 hPa, z is the altitude in meters, and H is a scale height of 7400 m.

The second quantity is the tropospheric vertical column density (tVCD), which is computed by summing the profile partial columns over all levels with a bottom pressure edge > 200 hPa. This is converted from Dobson Units to molec. cm^{-2} by multiplying by 2.69×10^{16} molec. cm^{-2} / DU.

Both the surface concentrations and tVCDs are averaged over all pixels binned to a given site for a given month. Unlike the NO_2 product, no weighting for pixel area is applied, as the pixel size is not given in this product.

2.3. Sonic mappings

After the preprocessing, we have a set of locations through the United States that each have corresponding time series for NO_2 tVCDs and O_3 concentrations. We present several modes for listening to the data. In the simplest case, we use the data for each location as the frequency parameter to a sinusoidal oscillator. We exponentially map the quantities of the compounds into an audible range according to

$$c \left[(d/c)^{\frac{x-a}{b-a}} \right] , \tag{2}$$

where x is the value being mapped from the old range $[a, b]$ to the new range $[c, d]$. In general, we can constrain NO_2 and O_3 to different frequency ranges. Moreover, we provide three scaling/normalization schemes for mapping the data for each compound at each site to frequency. In order to make comparisons across all sites, we scale the data by the global compound minimum and maximum. Since this compresses the frequency range of the individual sites, we can also scale the data by each site's individual minimum and maximum. While this makes it impossible to make judgments across sites, it highlights the seasonal and overall trends for individual sites. Last, we can scale the data by the minimum and maximum values of an individual type of location. This is particularly helpful for rural sites, as the quantities of NO_2 and O_3 change much less than in cities and power plants. In all cases, as the number of geographical locations increases, it becomes challenging to attend to which time series is which.

One spatial panning scheme presents each compound on independent speaker channels (e.g., $NO_2 \rightarrow$ left and $O_3 \rightarrow$ right). This makes it easier to identify the time series related to NO_2 and O_3. In order to improve the listener's ability to differentiate geographical locations, we also have a second panning scheme that maps each location's longitude and latitude to auditory spatial location. We place the listener in the center of the United States facing north. We offer a stereo mode, which maps longitude to left and right (i.e. the west coast of the United States to the left and the east coast to the right). An additional quadrophonic mode also maps latitude to front and back; this mode supposes speakers placed directly to the front, back, left, and right of the listener.

Another parameter of the sonification is the sound synthesis. Since Sine waves are challenging to localize in space, it can be hard to differentiate between simultaneously playing sites in the second panning mode. To address this issue, we propose several other sound generators to replace the sine oscillator. These include band-pass filtered noise, sawtooth oscillators, and square oscillators. These can be assigned to each site category; here, we use a square oscillator for cities, a sawtooth oscillator for power plants, and filtered noise for the rural sites.

One final sonification parameter is a "season filter" designed to highlight seasonal fluctuations. This is implemented as a resonant low-pass filter where the cutoff frequency oscillates as a function of season. For each site, the filter cutoff frequency is designed relative to that site's fundamental frequency. This filters the spectrum and acts as an auditory cue that highlights the effects of seasonal changes.

2.4. Sonification interface

While the synthesis routines do not depend on the GUI, having an interface makes it much easier to understand and control the parameters of the sonification. The GUI, which can be seen in Fig. 3, displays a map of the United State with indicators for where we present data. These locations are color coded by location type (e.g., power plants, cities, etc.). A bank of check-boxes allow the user to select which locations contribute to the sonification. Naturally, the map is updated to reflect which locations are selected.

Basic control is provided for playing/pausing and muting/unmuting. We also provide the ability to loop through the data or step through it one month at a time. Controls are provided to control the speed at which the sonification is performed. When the data are presented quickly, the listener perceives general trends while at slower speeds the seasonal and monthly trends dominate. This time resolution control is useful for zooming between micro and macro scales.

Finally, controls are offered for switching between the various sonification schemes. We provide the appropriate controls to enable/disable features of the model, and when appropriate, control their parameters.

3. EVALUATION

To evaluate the effectiveness of the sonification, informal listening tests were performed. Listeners primarily had backgrounds in either atmospheric chemistry or music, but not both. The listening tests involved a pretest to judge each listener's prior knowledge, followed by a guided and unguided exploration of the features and mappings of the sonification to gauge their effectiveness. We also present our own observations of trends that are evident in the data.

3.1. NO_2 interannual trends

We found NO_2 trends are apparent both for a single location and when comparing multiple locations. When using the global scaling, the seasonal oscillation in NO_2 tVCDs stands out clearly, especially for cities. For rural environments, the global scaling compresses the values. This is useful, however, for comparing their

absolute levels to urban locations. Aurally, it is obvious that the overall NO_2 levels are much lower in rural environments. Using the local scaling, the downward trends and seasonal oscillatory trends become more discernible.

In general, we found that power plants seem to be less affected by seasonal changes than cities, but listening to the locally scaled version of the power plants makes the seasonal changes more apparent. One interesting observation is that some power plants (e.g., Huntington) start with a sharp decrease in NO_2 emissions and then level out for the rest of the dataset. Overall, these observations provide compelling evidence for the importance of multiple scaling options.

We also observe that locations on the East Coast tend to have higher overall levels than elsewhere in the country. For example, Shenandoah is in rural Virginia, however the NO_2 levels are comparable to those in Fresno, California. Naturally, Shenandoah's NO_2 levels are influenced by its proximity to large cities like Baltimore. Likewise, trends in Philadelphia and Pittsburgh are highly correlated to one another.

While listening to all sites in a given category simultaneously, listeners were nearly all able to identify that both city and power plant NO_2 have markedly decreased between 2005 and 2015, while rural NO_2 remained constant. When asked to identify individual sites with the largest change in NO_2 tVCDs, listeners were generally able to correctly identify the city with the largest change out of four choices, although this may have been influenced by prior knowledge as very few listeners' confidence in this answer increased after listening. All listeners were able to identify at least one rural site with the largest change out of all fifteen sites, as long as they were permitted to turn individual sites on and off, and about half correctly identified the two rural sites with significantly larger decreases than the other thirteen rural sites. In both cases (comparing four cities and comparing all rural sites), listeners were not able to identify the which site had the largest change when multiple sites were playing simultaneously. Overall, we succeeded in our basic goal of representing decadal trends in NO_2 tVCDs.

3.2. O_3 interannual trends

For evaluation, we use the O_3 concentrations nearest the surface rather than the partial O_3 column. We observe that an overall rise in most cities is readily apparent, as are the seasonal oscillations, even for listeners without extensive knowledge of NO_x/O_3 chemistry.

This indicates that the OMO3PR product shows that O_3 is increasing on the interannual timescale throughout the US. This is consistent with the hypothesis that O_3 increases with temperature [34]. However, due to the difficulty in retrieving tropospheric O_3 from space, verifying this with other tropospheric O_3 satellite products (e.g. [35]) should be considered, and comparison to surface-based measurements should be carried out. Additionally, we are planning on including temperature data to demonstrate the relationship between O_3 and temperature. Nevertheless, the listener can clearly hear interannual trends in the O_3 retrieved with the OMO3PR product; thus the sonification itself represents this successfully and can be used with other O_3 measurements in the future.

Sonifying O_3 faces some challenges, given that the data has more extreme values compared to NO_2. These are comparatively infrequent, but significantly larger than the majority of values. In our global mapping, this means that the majority of the variability is compressed to a very small range. This can be alleviated by doing the same mapping as described in §2.3, but by clipping the range to prevent outliers from having such an effect on the rest of the data. This will make the interannual variability much clearer for most sites, but means the listener cannot compare absolute magnitudes. Given these difficulties and the relatively high uncertainty of the O_3 product, the informal listening tests did not test the perception of O_3 trends.

3.3. Panning, season filter, and synthesis sounds

In our listening tests, we were particularly interested in the success of these sonic mappings. As predicted by the resolution of human auditory spatial cues, listeners were better able to perceive and interpret longitudinal (left-right) spatial cues than latitudinal (front-back) spatial cues. It is challenging to mitigate this issue outright as we are fighting with a physiological feature of the human auditory system. Perhaps a higher speaker density and a spatially warped panning algorithm would improve the result of the spatial cues. Additionally, the longitudinal panning was insufficient to help most listeners identify individual sites when multiple sites were playing (§3.1). That being said, many listeners remarked that the panning improved their engagement throughout the listening test.

The season filter was unsuccessful for giving listeners seasonal cues. At high speeds, the cutoff oscillations are pronounced but the data are presented too quickly to make sense of the filtering. At slower speeds, the effects are masked. Both issues are accentuated because, while higher pitches connect very naturally to greater quantities of NO_2 or O_3, the cutoff frequency of a filter has no intrinsic connection to season. More problematically, some people remarked that the season filter affected their perception of the fundamental pitch used to convey the primary data. As this could confound the primary goal, we want to redesign the implementation of the season filter. One idea is to "detrend" the data when the season filter is engaged. By removing the overall (long time scale) trends, the localized fluctuation will hopefully become more pronounced. Another idea would be to allow the user to select which months to hear. Then the listener would not have to rely as much on memory. For instance, one could listen to only summer months to hear trends in that annual cycle rather than hearing the full year, or the playback could be set to advance in six-month steps, directly comparing summer and winter months.

The different sonic oscillators (filtered noise, sawtooth, etc) representing different site types were easy for some listeners to differentiate but challenging for others. When listening to multiple sites at once, listeners found it more challenging to differentiate between the square and sawtooth waveforms (cities and power plants) than the filtered noise (rural sites). Listeners also found it more challenging to discriminate pitch for the filtered noise.

3.4. Relationship between NO_2 and O_3

As described in §1.1, the changes in O_3 concentration can be positively or negatively correlated with changes in the NO_2 tVCDs. The informal listening tests had listeners attempt to identify whether NO_2 and O_3 were correlated, anti-correlated, or uncorrelated at Phoenix, AZ. Fig. 4 plots the change in NO_2 and O_3 between adjacent data-points as a function of time for Phoenix, AZ. The overall correlation between ΔNO_2 and ΔO_3 is negative (where Δ indicates a month to month difference), although there

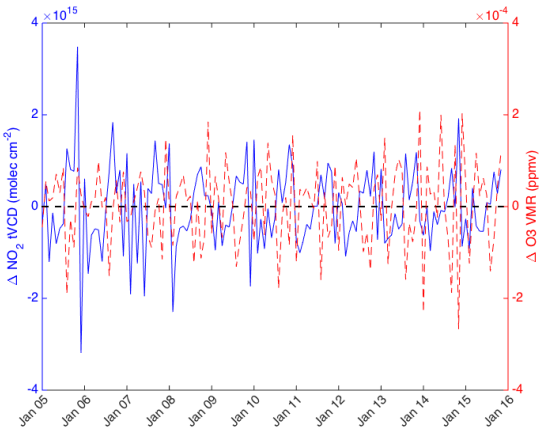


Figure 4: Delta-delta time-series plot of O_3 and NO_2 for Phoenix, AZ. Each point is the difference in NO_2 tVCD or O_3 concentration between that month and the following month.

are times when the correlation is positive. Statistically, there is a very weak negative overall correlation ($R^2 < 0.1$) between ΔO_3 and ΔNO_2 at Phoenix.

This commingling of positive and negative correlations makes discerning the correlative relationship between ΔNO_2 and ΔO_3 sonically difficult. An additional element of difficulty comes from the high uncertainty in the chosen O_3 satellite product. Future refinements will explore using alternate sources of O_3 data, including other satellite products and direct surface measurements of O_3.

For the purpose of evaluating the effectiveness of our sonification at communication the NO_2/O_3 relationship, we judge whether listeners identified the overall correlation: that NO_2 and O_3 are negatively correlated at Phoenix, AZ. Most listeners' response after listening moved towards anti-correlated compared to their pretest response. About half of listeners completely reversed their response after listening from correlated to anti-correlated. Therefore, we judge our sonification to be moderately successful at communicating the NO_2/O_3 relationship in this case.

4. CONCLUSIONS

In this paper, we have described a sonification tool for comparing trends in NO_2 tVCDs and O_3 concentrations across the United States. Overall, informal listening tests have shown that untrained listeners can readily identify interannual trends and identify which sites exhibit the largest changes in NO_2. Clearly, there is more work to be done. We need to provide additional data such as temperature and humidity as well as ground based measurements to improve the educational value of presenting the data as a sonification for untrained listeners. Certain secondary elements of the sonfication, particularly representing seasonal differences and the geographical position through panning need additional development to make them fully effective. Furthermore, while this project focuses on an experiential auditory experience for non-experts, we should provide more visual feedback (e.g., generating plots on-the-fly that correspond to the sonification) as an additional avenue for learning about NO_2 and O_3 trends.

5. ACKNOWLEDGMENT

JLL acknowledges support from NASA through the Earth and Space Science Fellowship NNX14AK89H, NASA grants NNX15AE37G and NNX14AH04G, and the TEMPO project grant SV3-83019. JLL also thanks Ronald C. Cohen for advice and support as a Ph.D. advisor. EKCD's research reported in this publication was partially supported by the National Academies Keck Futures Initiative of the National Academy of Sciences under award number NAKFI ADSEM10. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Academies Keck Futures Initiative or the National Academy of Sciences.

6. REFERENCES

[1] I. Romieu, F. Meneses, S. Ruiz, J. J. Sienra, J. Huerta, M. C. White, and R. A. Etzel, "Effects of air pollution on the respiratory health of asthmatic children living in Mexico City." *Am. J. Resp. Crit. Care*, vol. 154, no. 2, pp. 300–307, 1996.

[2] A. P. K. Tai, M. V. Martin, and C. L. Heald, "Threat to future global food security from climate change and ozone air pollution," *Nat. Clim. Change*, vol. 4, no. 9, pp. 817–821, 2014. [Online]. Available: http://dx.doi.org/10.1038/nclimate2317

[3] J. G. Murphy, D. A. Day, P. A. Cleary, P. J. Wooldridge, D. B. Millet, A. H. Goldstein, and R. C. Cohen, "The weekend effect within and downwind of Sacramento – Part 1: Observations of ozone, nitrogen oxides, and VOC reactivity," *Atmos. Chem. Phys.*, vol. 7, no. 20, pp. 5327–5339, 2007. [Online]. Available: http://www.atmos-chem-phys.net/7/5327/2007/

[4] P. Monks and S. Beirle, "Applications of Satellite Observations of Tropospheric Composition," in *The Remote Sensing of Tropospheric Composition from Space*, J. Burrows, U. Platt, and P. Borrell, Eds. New York: Springer, 2011.

[5] J. P. Burrows, M. Weber, M. Buchwitz, V. Rozanov, A. Ladstätter-Weißenmayer, A. Richter, R. DeBeek, R. Hoogen, K. Bramstedt, K.-U. Eichmann, M. Eisinger, and D. Perner, "The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results," *J. Atmos. Sci.*, vol. 56, no. 2, pp. 151–175, 1999. [Online]. Available: http://dx.doi.org/10.1175/1520-0469(1999)056⟨0151:TGOMEG⟩2.0.CO;2

[6] H. Bovensmann, J. P. Burrows, M. Buchwitz, J. Frerick, S. Nol, V. V. Rozanov, K. V. Chance, and A. P. H. Goede, "SCIAMACHY: Mission Objectives and Measurement Modes," *J. Atmos. Sci.*, vol. 56, no. 2, pp. 127–150, 1999. [Online]. Available: http://dx.doi.org/10.1175/1520-0469(1999)056⟨0127:SMOAMM⟩2.0.CO;2

[7] P. Levelt, G. van der Oord, M. Dobber, A. Mälkki, H. Visser, J. de Vries, P. Stammes, J. Lundell, and H. Saari, "The Ozone Monitoring Instrument," *IEEE Trans. Geosci. Remote Sense.*, vol. 44, pp. 1093–1101, 2006.

[8] J. Callies, E. Corpaccioli, A. Hahne, and A. Lefebvre, "GOME-2 – MetOp's Second-Generation Sensor for Operational Ozone Monitoring," *ESA Bulletin*, vol. 102, pp. 28–36, 2000. [Online]. Available: http://www.esa.int/esapub/bulletin/bullet102/Callies102.pdf

[9] J. Ding, R. J. van der A, B. Mijling, P. F. Levelt, and N. Hao, "NO_x emission estimates during the 2014 Youth Olympic Games in Nanjing," *Atmos. Chem. Phys.*, vol. 15, no. 16, pp. 9399–9412, 2015.

[10] L. N. Lamsal, B. N. Duncan, Y. Yoshida, N. A. Krotkov, K. E. Pickering, D. G. Streets, and Z. Lu, "U.S. NO_2 trends (2005–2013): EPA Air Quality System (AQS) data versus improved observations from the Ozone Monitoring Instrument (OMI)," *Atmos. Environ.*, vol. 110, pp. 130–143, 2015.

[11] D. Q. Tong, L. Lamsal, L. Pan, C. Ding, H. Kim, P. Lee, T. Chai, K. E. Pickering, and I. Stajner, "Long-term NOx trends over large cities in the United States during the great recession: Comparison of satellite retrievals, ground observations, and emission inventories," *Atmos. Environ.*, vol. 107, pp. 70–84, 2015.

[12] M. Huang, K. W. Bowman, G. R. Carmichael, T. Chai, R. B. Pierce, J. R. Worden, M. Luo, I. B. Pollack, T. B. Ryerson, J. B. Nowak, J. A. Neuman, J. M. Roberts, E. L. Atlas, and D. R. Blake, "Changes in nitrogen oxides emissions in California during 2005–2010 indicated from top-down and bottom-up emission estimates," *J. Geophys. Res. Atmos.*, vol. 119, pp. 12,928–12,952, 2014.

[13] G. C. M. Vinken, K. F. Boersma, A. van Donkelaar, and L. Zhang, "Constraints on ship NO_x emissions in Europe using GEOS-Chem and OMI satellite NO_2 observations," *Atmos. Chem. Phys.*, vol. 14, no. 3, pp. 1353–1369, 2014. [Online]. Available: http://www.atmos-chem-phys.net/14/1353/2014/

[14] D. Gu, Y. Wang, C. Smeltzer, and Z. Liu, "Reduction in NOx Emission Trends over China: Regional and Seasonal Variations," *Environ. Sci. Technol.*, vol. 47, no. 22, pp. 12 912–12 919, 2013.

[15] K. Miyazaki, H. Eskes, and K. Sudo, "Global NO_x emissions estimates derived from an assimilation of OMI tropospheric NO_2 columns," *Atmos. Chem. Phys.*, vol. 12, pp. 2263–2288, 2012.

[16] A. R. Russell, L. C. Valin, and R. C. Cohen, "Trends in OMI NO_2 observations over the United States: effects of emission control technology and the economic recession," *Atmos. Chem. Phys.*, vol. 12, pp. 12 197–12 209, 2012.

[17] J.-T. Lin, M. B. McElroy, and K. F. Boersma, "Constraint of anthropogenic NO_x emissions in China from different sectors: a new methodology using multiple satellite retrievals," *Atmos. Chem. Phys.*, vol. 10, pp. 63–78, 2010.

[18] S.-W. Kim, A. Heckel, G. J. Frost, A. Richter, J. Gleason, J. P. Burrows, S. McKeen, E.-Y. Hsie, C. Granier, and M. Trainer, "NO_2 columns in the western United States observed from space and simulated by a regional chemistry model and their implications for NO_x emissions," *J. Geophys. Res. Atmos.*, vol. 114, 2009.

[19] K. Miyazaki, H. Eskes, K. Sudo, and C. Zhang, "Global lightning NO_x production estimated by an assimilation of multiple satellite data sets," *Atmos. Chem. Phys.*, vol. 14, pp. 3277–3305, 2014.

[20] S. Beirle, H. Huntrieser, and T. Wagner, "Direct satellite observations of lightning-produced NO_x," *Atmos. Chem. Phys.*, vol. 10, pp. 10 965–10 986, 2010.

[21] P. Castellanos, K. F. Boersma, and G. R. van der Werf, "Satellite observations indicate substantial spatiotemporal variability in biomass burning NO_x emission factors for South America," *Atmos. Chem. Phys.*, vol. 14, pp. 3929–3943, 2014.

[22] A. Mebust and R. Cohen, "Space-based observations of fire NO_x emissions coefficients: a global biome-scale comparison," *Atmos. Chem. Phys.*, vol. 14, pp. 2509–2524, 2014.

[23] ——, "Observations of a seasonal cycle in NO_x emissions from fires in African woody savannas," *Geophys. Res. Lett.*, vol. 40, pp. 1451–1455, 2013.

[24] J. Zörner, M. Penning de Vries, S. Beirle, H. Sihler, P. R. Veres, J. Williams, and T. Wagner, "Multi-satellite sensor study on precipitation-induced emission pulses of NO_x from soils in semi-arid ecosystems," *Atmos. Chem. Phys.*, vol. 16, no. 14, pp. 9457–9487, 2016.

[25] R. Nassar, J. A. Logan, H. M. Worden, I. A. Megretskaia, K. W. Bowman, G. B. Osterman, A. M. Thompson, D. W. Tarasick, S. Austin, H. Claude, M. K. Dubey, W. K. Hocking, B. J. Johnson, E. Joseph, J. Merrill, G. A. Morris, M. Newchurch, S. J. Oltmans, F. Posny, F. J. Schmidlin, H. Vmel, D. N. Whiteman, and J. C. Witte, "Validation of Tropospheric Emission Spectrometer (TES) nadir ozone profiles using ozonesonde measurements," *J. Geophys. Res. Atmos.*, vol. 113, no. D15, 2008, d15S17. [Online]. Available: http://dx.doi.org/10.1029/2007JD008819

[26] United States Environmental Protection Agency. (1999) Nitrogen Oxides (NOx), Why and How They Are Controlled (EPA-456/F-99-006R). [Online]. Available: https://www3.epa.gov/ttncatc1/dir1/fnoxdoc.pdf

[27] ——. (2016) Transportation Air Quality, Selected Facts and Figures. [Online]. Available: https://www.fhwa.dot.gov/environment/air_quality/publications/fact_book/factbook2016.pdf

[28] ——. Clean Air Interstate Rule (CAIR). [Online]. Available: https://archive.epa.gov/airmarkets/programs/cair/web/html/index.html

[29] Z. Lu, D. Streets, B. de Foy, L. Lamsal, B. Duncan, and J. Xing, "Emissions of nitrogen oxides from US urban areas: estimation from Ozone Monitoring Instrument retrievals for 2005–2014," *Atmos. Chem. Phys.*, vol. 15, pp. 10 367–10 383, 2015.

[30] M. Kroon, J. F. de Haan, J. P. Veefkind, L. Froidevaux, R. Wang, R. Kivi, and J. J. Hakkarainen, "Validation of operational ozone profiles from the Ozone Monitoring Instrument," *J. Geophys. Res.*, vol. 116, no. D18, 2011. [Online]. Available: http://dx.doi.org/10.1029/2010JD015100

[31] J. de Haan and J. Veefkind. (28 Sept 2012) OMO3PR Readme. [Online]. Available: https://disc.gsfc.nasa.gov/Aura/data-holdings/OMI/documents/v003/OMO3PRO_README.shtml

[32] KNMI. (26 Oct 2012) Background information about the row anomaly in omi. [Online]. Available: http://projects.knmi.nl/omi/research/product/rowanomaly-background.php

[33] National Oceanic and Atmospheric Administration. The Global Land One-km Base Elevation Project. [Online]. Available: https://www.ngdc.noaa.gov/mgg/topo/globe.html

[34] M. Lin, L. W. Horowitz, R. Payton, A. M. Fiore, and G. Tonnesen, "US surface ozone trends and extremes from 1980 to 2014: quantifying the roles of rising Asian emissions, domestic controls, wildfires, and climate," *Atmos. Chem. Phys.*, vol. 17, no. 4, pp. 2943–2970, 2017. [Online]. Available: http://www.atmos-chem-phys.net/17/2943/2017/

[35] Y. Choi, Y. Wang, Q. Yang, D. Cunnold, T. Zeng, C. Shim, M. Luo, A. Eldering, E. Bucsela, and J. Gleason, "Spring to summer northward migration of high O3 over the western North Atlantic," *Geophys. Res. Lett.*, vol. 35, no. 4, 2008. [Online]. Available: http://dx.doi.org/10.1029/2007GL032276

SOLAR SYSTEM SONIFICATION: EXPLORING EARTH AND ITS NEIGHBORS THROUGH SOUND

Brianna J. Tomlinson[1], R. Michael Winters[2], Christopher Latina[2], Smruthi Bhat[3], Milap Rane[2], Bruce N. Walker[1,4]

School of Interactive Computing[1],
Georgia Tech Center for Music Technology (GTCMT)[2],
School of Computer Science[3], School of Psychology[4]
Georgia Institute of Technology, Atlanta, USA
btomlin@gatech.edu, mikewinters@gatech.edu, chrisrlatina@gmail.com,
smruthib@gatech.edu, mrane3@gatech.edu, bruce.walker@psych.gatech.edu

ABSTRACT

Informal learning environments (ILEs) like museums incorporate multi-modal displays into their exhibits as a way to engage a wider group of visitors, often relying on tactile, audio, and visual means to accomplish this. Planetariums, however, represent one type of ILE where a single, highly visual presentation modality is used to entertain, inform, and engage a large group of users in a passive viewing experience. Recently, auditory displays have been used as a supplement or even an alternative to visual presentation of astronomy concepts, though there has been little evaluation of those displays. Here, we designed an auditory model of the solar system and created a planetarium show, which was later presented at a local science center. Attendees evaluated the performance on helpfulness, interest, pleasantness, understandability, and relatability of the sounds mappings. Overall, attendees rated the solar system and planetary details very highly, in addition to providing open-ended responses about their entire experience.

1. INTRODUCTION

Museums and other ILEs have explored multi-modal exhibits to increase engagement and prolong interaction for attendees [1, 2, 3]. Multi-modal exhibit design provides additional methods for presenting content to visitors, enhancing the experience for everyone, and allowing greater access to those with impairments [4]. Larger movements in incorporating universal design have resulted in greater development and evaluation of accessibility in these learning environments [5]. Descriptive audio tours and other auditory displays can support shared experiences for larger groups of visitors [6] and provide exploration methods for those with vision impairment [7, 8, 9]. One example, the Aquarium Sonification created dynamic soundscapes through mapping fish characteristics and events within the tank, as a way to provide a unique experience for individual exhibits [3]. On the other hand, not all ILEs use multiple modalities in their presentations. As one example, planetarium shows are typically visual-only; or, if there is any audio it is largely supplementary.

This paper explores and expands the use of these auditory experiences in a planetarium show—a typically informal, passive learning experience that is predominantly visual. Leveraging the possibilities of spatial audio, a variety of quantitative information about each planet in our solar system was conveyed to an audience through sound (with some static visual anchors). Surveys collected during the show demonstrate that the show was interesting, understandable, relatable and helpful, even to a sample audience without visual impairments. The results hold promise for the creation of future shows that entertain and educate through listening.

2. RELATED WORK

Sonification, or the use of non-speech sound to present information, has been explored and used for a variety of situations and applications as a type of auditory display [10]. Even though there is some precedence for using auditory displays as a way to promote public interest in space and astronomy, there has not been extensive evaluation of these displays. Previous work has focused on making sounds that are already collected through (radio) telescopes and other instruments available to the public. Harger and Hyde, and others have broadcast live sounds from radio telescopes over the internet and FM radio stations [11, 12].

Some work has explored using sonification and audification (direct mapping of a dataset to sounds) to analyze data sets from space, such as Cosmic Microwave Background Radiation or the Search for Extra-Terrestrial Intelligence (SETI) [13]. Other recent work includes Landi et al.'s analysis of solar wind through audification of solar rotation data to explore carbon ionization [14]. Ballora created more musically-composed sonifications for an outreach film presented at the Smithsonian Air & Space Museum, but did not evaluate their success in presenting the information to attendees [15].

Recently, Quinton et al. developed a model for representing characteristics of the Solar System [16]. Through an interview with a planetarium representative, they identified seven important concepts to include in their model (density, diameter, gravity, length of day, orbital period, temperature, and orbital velocity). They completed a small-scale evaluation with 12 users, where each participant was asked to identify characteristics for each planet when listening to the sonification. Though they gathered valuable

https://doi.org/10.21785/icad2017.027

feedback about their sonifications, this work was focused more on individual interpretation of the model without scaffolding their experience and knowledge and did not evaluate the work in an ecologically valid manner.

Exploring how to represent planetary data in a multi-modal experience and evaluating different characteristics of the design for the solar system model presents a novel, artistic, and interesting area to explore, which has the potential to reach into other applications in both informal learning spaces and formal learning activities.

3. IDENTIFYING IMPORTANT CONCEPTS

At the beginning of the development, we wanted to better understand which concepts were most important for teaching a comprehensive understanding about the solar system. To do this, we conducted semi-structured interviews with five teachers across different levels of science classes, including a planetarium instructor, a university professor who teaches intro level astronomy, and three additional teachers across elementary through high school. All of the teachers had at least 9 years of teaching experience, with a few having almost 30 years of experience teaching astronomy.

During the interviews, we asked about the types of space-related concepts they teach, some detailed examples of how they introduce the topics, and the types of misconceptions that confuse their students (or audience) the most. We used these interviews as a way to explore how to structure the introduction of the topics in the planetarium show and a way to identify which information people struggle with learning the most.

A common theme that arose from these interviews was the lack of prior knowledge most people have about astronomy: the teachers need to start their lessons by introducing everything from the ground up. Many of the teachers described comparing and contrasting features of the planets such as the gravitational strength, atmosphere and surface composition, and other details such as rings and moons. The planetarium instructor mentioned mixing different levels of detail for each of the topics, in case some individuals in her audience already know the basics. This is more of a concern for her than it is for the other teachers, who typically have more homogeneous groups of learners.

When asked about common misconceptions their students have, all of the teachers said that understanding the scale of the solar system (and space in general) was one of the hardest concepts to convey. Another common misconception dealt with understanding seasons (and their relation to the tilt of a planet) and the elliptical nature of orbits.

Using the interview responses from the teachers, we decided to focus our sonification on details related to the scale of the solar system, such as mass, temperature, and distance.

4. SOUND DESIGN

The process for designing the sounds for the show was focused around two "views" of the solar system. In the first view, the Solar System View, the planets' masses, lengths of years, lengths of days, and distances from the sun were compared to each other, first as if the listeners were on the sun, hearing them rotate around; and then as if traveling on a space-ship visiting each planet on the way. This view worked to address the misconception of size and scale identified through the interviews, and presented a baseline of information about the planets for everyone in the audience.

In the second view, the Planetary View, the desire was to provide an experience of what it would be like to be on a given planet's surface. In this view, each planet's type, number of moons, number of rings, gravitational force, and temperature range were conveyed. This view provided additional details about the features of the planets and allowed for comparisons between them.

For all of the cases, sound was spatialized using the planetarium's quadraphonic speaker system and Vector-Based Amplitude Panning (VBAP). This configuration allowed sources to rotate around the audience at variable speeds. For example, in the Solar System View, the sounds representing the planets were first distributed spatially, and rotated around the listener at a rate proportional to the orbital period of that planet around the sun. Two speed factors were introduced to make the planet's relative motions fast enough to be perceived by ear, one factor was applied to the terrestrial planets and one factor for the gas giants (which are much slower). For the Planetary View, a similar approach was taken for moons and rings, though without taking into account the specific orbital periods of individual moons. The choice of a quadraphonic speaker algorithm was guided by the affordances of the planetarium's audio system, which already included four equidistant speakers positioned in a ring around the audience area.

Prototyping and designing the sonification made use of SuperCollider [17, 18], an open-source computer music library commonly used in sonification [19]. Data for sonification were downloaded from NASA's Planetary Fact Sheet [20], and mapping principles were derived from principles of stream segregation and parameter nesting, as well as designing to support spatial audio [21, 22, 23, 24].

4.1. Solar System View

The Solar System View individually conveyed the mass of each planet, the length of year, the length of day, and the distance of each planet from the sun. To create the fundamental sound of each planet, brown noise was used, and a resonant filter was applied whose center frequency scaled proportionally to the mass of the planet. Following the polarity mappings of previous work [25], the mass of each planet was mapped inversely to pitch (i.e., sounds for larger planets had a lower pitch).

To encode the length of day, we modulated the amplitude envelope of the filtered noise sinusoidally between zero amplitude and full amplitude. The frequency of this modulation was linearly proportional to the length of the day. The analogy of this strategy was the cycle of sunrise and sunset on a planet: a day would be perceived as a gradual increase in lightness from the sun (increasing volume), and decay with the end of the day (decreasing volume). The planets vary tremendously in speed of rotation around their axis. However, a constant scaling factor was applied that allowed most planets to fall in the range of human rhythm cognition [26], leaving only a few planets (e.g., Mercury, Venus), whose length of day were too long for the percept of a tactus (recognizable beat) to form. The transformation we used converted 24 hours (1 day on earth) into 1 second.

The length of year was conveyed using spatial location and speed of revolution around the listener. The shorter the length of year, the faster it would revolve around the listener. Due to the tremendous range of the length of years in the planets, these were conveyed in two phases. One phase for the four terrestrial planets, which have shorter years, and a second for the four gas giants, which have longer years. Jupiter was used as a reference to link

between the two phases. In the first phase, it moved the slowest. In the second phase, it moved the fastest.

To convey distance from the sun, a spaceship traveled to each planet in sequence starting at the Sun and finishing on Neptune. Although there is no sound in space, a fictitious sound effect was created for the spacecraft with additional sound effects for passing objects like asteroids. This base sound was used to convey the distance between the first four terrestrial planets, which would be 3-4 seconds apart. To get to the gas giants, the ship was "accelerated" to five times its normal speed to reach Jupiter and Saturn, and accelerated to ten times the normal speed to reach Uranus and Neptune. The acceleration and velocity were conveyed by increasing the playback speed of the spaceship sound proportionally to the new speed, meaning a faster spaceship speed had a higher pitch. This acceleration had the affect of allowing the ship to reach the planets within a reasonable time-scale in the show.

4.2. Planetary View

Based on our interviews with astronomy educators, we chose specific features to cover for each of the planets: moons, rings, temperature range, gravitational strength, and type of planet. We mapped the number of moons to an equal number of high-frequency sinusoids rotating at variable speeds around the listener. The frequency of each moon was within a 3 octave range above C5 and varied in loudness, but unlike in the solar system view, the precise mass and rotational period of each moon was not used in the mapping. Frequencies were randomly selected with equal probability within the range. Using sinusoids instead of brown-noise gave the moons a bell-like sound and also increased the differentiability of each moon. This choice allowed more moons to be conveyed in a narrower high-frequency space than if we had used filtered noise, gracefully handling the difference between representing just Earth's moon versus over 60 belonging to Jupiter and Saturn.

We encoded the number of rings through a distribution of variable pitch and loudness pure-tones as well, but instead of having specific spatial locations, they were played with equal amplitude through all the speakers. Additionally, a larger frequency range was used (extending downwards to C3) meaning that a ring could be conveyed using a much lower pure tone than a moon. Each pure-tone represented one ring, making scalability of number of rings easily portrayed.

An auditory graphing approach was used to display the temperature range of each planet whereby cold temperature was conveyed by low pitch and the highest temperature was conveyed by high pitch. Temperatures were first normalized across the planets and then the mean normalized temperature was multiplied by 1500 Hz with 200Hz added to each. The pitch range was determined by multiplying the normalized temperature range of each planet by 1500Hz. Short, 10Hz sine-grains were used to convey the temperature progression, which would increase and decrease at a rate proportional to the length of day, conveying that the temperature was a function of length of day.

The effect of gravity is proportional to the mass of each planet, and another sonification approach was used to convey the magnitude of the force of gravity. Using a physical model of a bouncing object, the effect of gravity was manipulated according to the force of gravity on the planet. As with the original mass mapping, the pitch of the sound of the ball hitting the surface was proportional to the size of the planet, so larger planets had a lower pitch than smaller planets.

The surface characteristics of the planets can be categorized into two broad types: terrestrial and gas giant. We designed a continuous timbre space to convey each type, which depended upon planetary mass and density. Beginning with the previously discussed mass-dependent fundamental frequency, a variable number of harmonics were added, the number of which increased logarithmically with mass. To convey density, the sound was passed through a low-pass filter whose cutoff frequency decreased with increasing density. This design gave the more diffuse gas giants a lower but richer timbre than the terrestrial planets, who were higher in pitch and more mellow in timbre.

4.3. Script Description

Although sonification was used to convey the information of the planets, speech was used to provide a sense of narrative to the show, explain to the audience what was being heard, and provide additional context. The script was written and recorded in a sound-isolated recording room and included two male narrators. The script had four sections: an introduction, the Solar System View, the Planetary View, and a conclusion. The first three portions of the show were used to present the different mappings and various facts about the planets, and the show concluded with an independent musical composition chosen for the show.

5. PLANETARIUM INSTALLATION

5.1. Location and Setup

After completing the design of the auditory display mappings, recordings, and the accompanying speech descriptions, we presented the show in a local planetarium at the end of April 2016. This event was free and open to the public. Participants were informed about the show through emails to campus mailing lists, posters, and advertisements posted at the science center in the weeks up to the show. While there is some potential for recruitment bias, the attendees (individuals who go out of their way to attend an ILE experience on a weeknight) are actually very representative of a typical evening show attendee.

The show leveraged the planetariums' quadraphonic speaker system for the spatial audio mappings, and prior to the show the gain of each speaker was referenced across frequencies to check for anomalies in the frequency response. As a way to supplement the sonifications and the descriptive script for the show, we created a PowerPoint presentation which contained images of each planet to be projected onto the dome as visual anchors for each of the sonifications. Using the four built-in projectors for the planetarium, this slide show accompanied the sonification.

As attendees entered the planetarium, we informed them about the exploratory nature of this research and asked if they wanted to provide feedback through a short paper survey at two points during the show. For attendees who agreed, they were given the survey that was filled out on two occasions, as follows.

For the first part of the show, attendees listened to a brief introduction and then the Solar System View, which included sonifications for the mass of each planet, the length of day, length of year, and distance from the sun. These were presented in groups, starting with Mercury and moving out toward Neptune (i.e., the mass for each planet was played, then the length of day, etc.). After the first part was completed, we paused the show and asked the

audience to complete a short survey of five six-point Likert-type questions regarding information mappings from the past series of sounds. As the audience was not able to individually explore the sounds at their own pace, we did not ask about individual sound mappings for each planet, but instead focused on the overall coherency and understandability of the sounds as a whole.

After the break for the first half of the survey, we resumed the second portion of the show. It contained two parts: the Planetary View and the conclusion. The Planetary View included separate sections for each planet's number of moons, presence (and number) of rings, overall temperature range, type of planet, and a representation of gravitational strength. Using sonifications in the planetary view allowed us to make comparisons which someone might not be able to easily make comparing static images of the planets. One example is how Venus and Uranus have very similar gravitational strength, though just observing them would not lead to this conclusion. During the show, their gravities (represented by the ball bouncing metaphor) were sequentially played making them easier to compare through other means. Similar comparisons were made between Uranus and Neptune, which have very similar characteristics (temperature and composition).

The final portion of the show was an artistic composition that recalled different auditory displays from the earlier, introductory portions of the show. At the end of the performance, the attendees completed the second set of Likert-type questions (about the planetary view mappings) and a series of free response questions regarding their overall listening experience.

5.2. Survey Description

The survey included two sections of Likert-type questions (one set for the Solar System View and one for the Planetary View), where the attendees were asked to rate the sounds on their aesthetics, helpfulness, and relatability to astronomical concepts:

1. How interesting were the sounds?
2. How pleasant were the sounds?
3. How helpful were the sounds?
4. How easy was it to understand the sounds?
5. How relatable were the sounds to their ideas?

We chose to use a six-point scale instead of a more typical five or seven-point scale to let the attendees give a range of either positive or negative feedback. Measuring the user experience for a listener can be difficult and sound often evokes diverse feelings from all individuals. The even scale was used to encourage the audience to be honest about their opinions through introspective reflection instead of choosing the (possibly) easier neutral option. For each question, one represented the lowest rating and six represented the highest.

The free response questions asked about overall likes and dislikes of the show, and asked about the attendee's favorite sound or set of sounds. We explored how well the audience members thought the planets were represented based on their previous knowledge. We asked the audience about how the sonifications made them feel, and whether or not they learned something interesting about the solar system or now appreciate another detail they did not know previously. Finally, we asked them if any portion of their understanding of the solar system has changed.

We asked two demographics questions, as the audience was open to all members of the community: age and student status (if a student, we also asked for current year in school).

6. RESULTS

Over 50 attendees came the night of the planetarium show, and 40 people completed the survey providing feedback about their experience throughout and after the show. Those who completed the survey reported their ages to be between 11 and 63, with the majority being in their 20s. Nineteen reported being students (ranging from fifth grade through graduate school). The attendees were asked to rate the sonifications to give high-level feedback on the sounds for the Solar System View and the Planetary View. As this was an initial exploration of a sonification model of the solar system, we did not perform inferential statistical analyses, but instead focused on descriptive statistics such as average ratings and the standard deviations (SD) of those ratings.

6.1. Quantitative Feedback

During the show, attendees responded to two categories of questions regarding the aesthetics (interesting and pleasant) and usefulness (helpful, understandable, and relatable) of the sonifications for the Solar System and Planetary Views. Table 1 presents the average rating for each question, while the detailed distributions can be found in Figure 1, Figure 2, and Figure 3. Each of the five categories achieved high ratings (at least 4.3 out of 6). The audience rated the aesthetics of the sounds especially highly (at least a 4.7 out of 6), with one category (Planetary View-interesting) having a 5 out of 6. Our carefully designed model of the solar system and information mappings was found tasteful and fitting by the attendees.

Criteria	Solar System View		Planetary View	
Interesting?	4.83	(1.11)	5.00	(0.95)
Pleasant?	4.73	(0.96)	4.84	(0.92)
Helpful?	4.68	(0.73)	4.55	(0.89)
Understandable?	4.75	(0.90)	4.31	(0.86)
Relatable?	4.59	(0.91)	4.49	(1.02)

Table 1: On a scale of 1 to 6, the users gave average ratings listed (and standard deviations).

The second grouping of questions sought to measure the audience's perceived usefulness of the auditory display and the overall understandability of the sonification mappings. Overall, the audience rated these three categories for both views quite highly, though the Solar System View had slightly higher scores. Ease of understanding for the Planetary View proved most difficult for the audience, perhaps due to each planet containing up to six different sonifications; across eight planets, this could lead to some confusion.

The most agreement between ratings (i.e., the responses that had the lowest variability) were for the helpfulness of the sonifications in understanding the concepts presented for both the Solar System and Planetary Views. The qualitative feedback from the audience in the next section provides some examples of how the sonifications were helpful for increasing their understanding of our solar system.

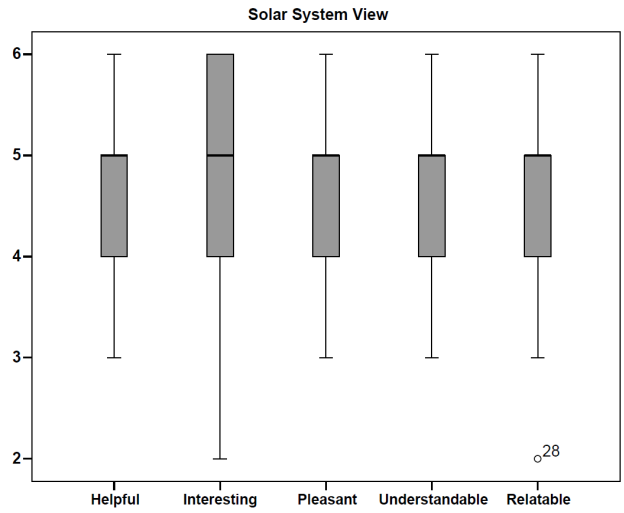


Figure 1: A series of boxplots representing the variation of the Solar System View Likert responses.

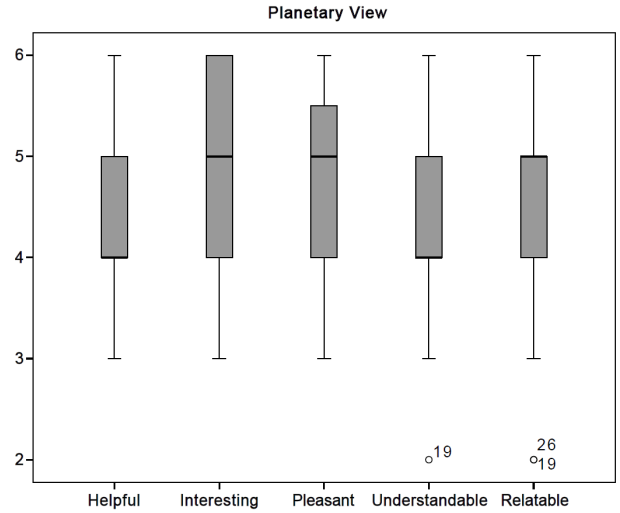


Figure 2: A series of boxplots representing the variation of the Planetary View Likert responses.

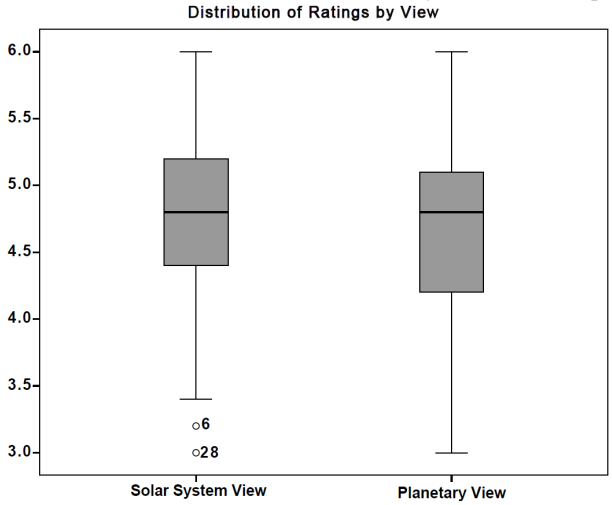


Figure 3: Two boxplots representing the overall variation range for both views.

6.2. Qualitative Feedback

In addition to the Likert-type ratings, we asked free response questions at the end of the show. These questions included listener likes or dislikes, favorite sounds, if there was something new they learned, and if they had any affect (emotional reaction) when listening to the solar system sonification. Many of the attendees reported particularly enjoying the sounds for the gas giants, the sounds of planets orbiting around them, the mapping for gravity (the ball bounce model), and they really enjoyed the portions that compared different planets to each other across one or two variables such as temperature, size, and distance from the sun.

Some of the respondents reported that it was much harder to remember the sounds during the second portion of the show, where we individually introduced information about each of the planets. There were many more details to remember in this section than in the first portion, which might explain the slightly lower averages in the second half of the show. However, even with those ratings, many people stated that they better understood the scale and the relationship between the planets; these comparisons were presented in the planetary view. Eighteen of the 40 survey respondents (45%) listed at least one piece of information they learned during the show that they had not known before, with those responses being evenly split between the first and the second half.

When asked if their understanding of the solar system changed (and how), one attendee reported, “Yes, the relationship between the planets is clearer” and another mentioned that they “realized the vast differences [between the planets]” after the show. Ten respondents provided comments similar to one attendee, who remarked that “The sound really helped put the distance in perspective,” referring to the first half of the show, and that “relative differences were very evident and cool to listen to.” During the interviews at the beginning of this research, the astronomy instructors mentioned one of the hardest concepts for students to understand

is the size and scale of the solar system. The sonifications helped the audience to understand them in a way they had not previously.

Many of the attendees recognized comparisons they had not previously noticed such as how the "weather/atmosphere comparisons with rings and moons was really interesting and informative." When respondents listed the details about the solar system they learned during the planetarium show, they reported that they had a better understanding of the relationships and comparisons between planets, especially gravity.

Finally, we asked all of the attendees to provide feedback about their overall experience during the show and we prompted them to explain how listening to the planets made them feel. Nineteen out of the 27 respondents who answered this question (70%) specifically stated how happy, joyful, or otherwise positive they felt from listening to the sonifications. One of the attendees explained how their mood changed throughout the show:

> *At first it was really overwhelming, almost headache inducing; however, the more the sounds were explained and pulled apart, the more meaningful and enjoyable it became.*

A few individuals reported how they felt "unexpectedly" relaxed while listening to the sonifications, and many reported how "pleasant and interesting" the experience was. For others, the sounds triggered strong aesthetic experiences, reporting that it made them feel "small and insignificant," while others more positively reflected that the sounds made them feel "awesome and ethereal."

There were no real differences between the ratings from the students based on grade, though the students in the lowest grades gave lower overall ratings to the five criteria than the older students did. Additionally, those attendees who did not mention deeper understanding of the solar system or learning new information had much lower ratings for the sonifications than those who did. Exploring these lower ratings more in-depth would be important before future deployment of the solar system sonifications in any other informal or formal learning context.

7. DISCUSSION

Overall, the audience gave each of the sonifications high ratings on average, with the highest scores from respondents on the aesthetic design of the solar system model. The well-designed sonifications provided an enjoyable experience to all audience members. Understanding and relating the sounds to their ideas proved most complex for the audience, and using their feedback to iterate on the model will decrease complexity and increase usefulness in future deployments.

The Planetary View presented many more individual details, and the attendees reported having more trouble remembering the sonifications for each of them. A way to support better learning and understanding of those portions of the sonification could be through an interactive exhibit, where the user could individually control and explore the sounds at their own pace. For future planetarium shows, we might also select a few representative cases for comparisons instead of presenting each interesting comparison available. Another possibility includes using smaller sections for each data characteristic instead of combining characteristics for one planet into a single section (presenting gravity, temperature, etc. in groups instead of every detail for one planet then the next).

Future work should also include evaluating the sounds' abilities to support learning this content directly. The evaluation for this show focused on exploring the audience's reaction and their experience attending a multi-modal ILE sonification. Additional deployments using the sonification as a show experience for student field trips would fit with the other major use case of the planetarium, and would present an opportunity to evaluate the sonification model in a way which might reduce potential confounds from demand characteristics (i.e., the students might be asked to give feedback on the show from the normal planetarium instructor, and not have biases from the researcher's introduction).

8. CONCLUSION

Planetariums typically rely on visuals (with varying levels of speech descriptions), but have not explored using sonifications or other auditory displays as a way to present information about space. Our positive results with respect to helpfulness, interest, pleasantness, understandability, and relatability of concepts demonstrate that attendees at a planetarium show can enjoy and learn information about space and our solar system by listening. In particular, the mappings and sound design used in our model of the solar system were successful, and future applications could involve a much wider audience, including visitors with vision impairment. Further study of the sonifications in a more traditional classroom context could also provide an interesting way to engage a diverse group of students through a multi-modal experience.

9. ACKNOWLEDGMENTS

Portions of this work were supported by funding from the National Science Foundation (NSF) and from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR). Thank you to the local science center for supporting and encouraging this work.

10. REFERENCES

[1] S. Allen, "Designs for learning: Studying science museum exhibits that do more than entertain," *Science Education*, vol. 88, no. S1, pp. 17–33, 2004, http://doi.org/10.1002/sce.20016.

[2] M. S. Horn, E. T. Solovey, R. J. Crouser, and R. J. Jacob, "Comparing the use of tangible and graphical programming languages for informal science education," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 975–984, http://dl.acm.org/citation.cfm?id=1518851.

[3] B. N. Walker, M. T. Godfrey, J. E. Orlosky, C. Bruce, and J. Sanford, "Aquarium sonification: Soundscapes for accessible dynamic informal learning environments." 12th International Conference on Auditory Display, ICAD 2006, 2006, https://smartech.gatech.edu/handle/1853/50426.

[4] Z. Obrenovic, J. Abascal, and D. Starcevic, "Universal accessibility as a multimodal design issue," *Communications of the ACM*, vol. 50, no. 5, pp. 83–88, 2007, http://www.realtechsupport.org/UB/I2C/UniversalAccessibility_2007.pdf.

[5] A. Elliott, "Developing accessible museum curriculum: the research, development and validation of a handbook

for museum professionals and educators," Ph.D. dissertation, Kansas State University, 2007, http://krex.k-state.edu/dspace/handle/2097/295.

[6] R. E. Grinter, P. M. Aoki, M. H. Szymanski, J. D. Thornton, A. Woodruff, and A. Hurst, "Revisiting the visit:: understanding how technology can shape the museum visit," in *Proceedings of the 2002 ACM conference on Computer supported cooperative work*. ACM, 2002, pp. 146–155, http://dl.acm.org/citation.cfm?id=587100.

[7] G. Anagnostakis, M. Antoniou, E. Kardamitsi, T. Sachinidis, P. Koutsabasis, M. Stavrakis, S. Vosinakis, and D. Zissis, "Accessible museum collections for the visually impaired: combining tactile exploration, audio descriptions and mobile gestures," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. ACM, 2016, pp. 1021–1025, http://doi.org/10.1145/2957265.2963118.

[8] A. Dulyan and E. Edmonds, "Auxie: initial evaluation of a blind-accessible virtual museum tour," in *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. ACM, 2010, pp. 272–275, http://dl.acm.org/citation.cfm?id=1952280.

[9] M. Jeon, R. J. Winton, J.-B. Yim, C. M. Bruce, and B. N. Walker, "Aquarium fugue: interactive sonification for children and visually impaired audience in informal learning environments." 18th International Conference on Auditory Display, ICAD 2012, 2012, https://smartech.gatech.edu/handle/1853/44427.

[10] B. N. Walker and M. A. Nees, "Theory of sonification," in *The sonification handbook*, A. H. Thomas Hermann and J. G. Neuhoff, Eds. Berlin: Logos Publishing House, 2011, pp. 9–39.

[11] J. C. Ballesteros and B. Luque Serrano, "Using sounds and sonifications for astronomy outreach," 2008, http://oa.upm.es/4724/.

[12] H. Harger and A. Hyde, "Broadcasting the music of the spheres: Creating radio astronomy," in *55th International Astronautical Congress*, 2004, http://doi.org/10.2514/6.IAC-04-IAA.6.16.1.04.

[13] P. Lunn and A. Hunt, "Listening to the invisible: Sonification as a tool for astronomical discovery," 2011, http://eprints.hud.ac.uk/15922.

[14] E. Landi, R. Alexander, J. Gruesbeck, J. Gilbert, S. T. Lepri, W. Manchester, and T. H. Zurbuchen, "Carbon ionization stages as a diagnostic of the solar wind," *The Astrophysical Journal*, vol. 744, no. 2, p. 100, 2011, http://dx.doi.org/10.1088/0004-637X/744/2/100.

[15] M. Ballora, "Sonification strategies for the film rhythms of the universe." 20th International Conference on Auditory Display, ICAD 2014, 2014, https://smartech.gatech.edu/handle/1853/52075.

[16] M. Quinton, I. McGregor, and D. Benyon, "Sonifying the solar system," 2016, http://www.icad.org/icad2016/proceedings2/papers/ICAD2016_paper_3.pdf.

[17] J. McCartney, "Rethinking the computer music language: Supercollider," *Computer Music Journal*, vol. 26, no. 4, pp. 61–68, 2002.

[18] S. Wilson, D. Cottle, and N. Collins, *The SuperCollider Book*. The MIT Press, 2011.

[19] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," *PloS one*, vol. 8, no. 12, p. e82491, 2013.

[20] D. R. Williams, "Planetary fact sheet," 2015, [Online; accessed 18-February-2017]. [Online]. Available: http://nssdc.gsfc.nasa.gov/planetary/factsheet/

[21] A. S. Bregman, "Auditory scene analysis: Hearing in complex environments," in *Thinking in sound: The cognitive psychology of human audition*. Cambridge: MIT Press, 1993, pp. 10–36.

[22] G. Kramer, "Some organizing principles for representing data with sound," in *Santa Fe Institute Studies in the Science of Complexity - Proceedings Volume*, vol. 18. Addison-Wesley Publishing Co, 1994, pp. 185–185.

[23] J. H. Schuett and B. N. Walker, "Measuring comprehension in sonification tasks that have multiple data streams," in *Proceedings of the 8th Audio Mostly Conference*. ACM, 2013, p. 11, http://doi.org/10.1145/2544114.2544121.

[24] J. H. Schuett, R. J. Winton, J. M. Batterman, and B. N. Walker, "Auditory weather reports: demonstrating listener comprehension of five concurrent variables," in *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*. ACM, 2014, p. 17, http://doi.org/10.1145/2636879.2636898.

[25] B. N. Walker, "Magnitude estimation of conceptual data dimensions for use in sonification," *Journal of Experimental Psychology*, vol. 8, no. 4, pp. 211–221, 2002, http://doi.org/10.1037/1076-898X.8.4.211.

[26] C. L. Krumhansl, "Rhythm and pitch in music cognition." *Psychological bulletin*, vol. 126, no. 1, p. 159, 2000.

MADBPM: MUSICAL AND AUDITORY DISPLAY FOR BIOLOGICAL PREDICTIVE MODELING

K. Michael Fox, Jeremy Stewart, Rob Hamilton

Rensselaer Polytechnic Institute
Department of the Arts
110 Eighth Street
Troy, New York 12180
{foxk, stewaj5, hamilr4}@rpi.edu

ABSTRACT

The modeling of biological data can be carried out using structured sound and musical process in conjunction with integrated visualizations. With a future goal of improving the speed and accuracy of techniques currently in use for the production of synthetic high value chemicals through the greater understanding of data sets, the *madBPM* project couples real-time audio synthesis and visual rendering with a highly flexible data-ingestion engine. Each component of the *madBPM* system is modular, allowing for customization of audio, visual and data-based processing.

1. INTRODUCTION

Data sonification is a rapidly evolving discipline that explores the use of data representing information from scientific processes, analytic models or real-time tracking as the drivers for sound generating systems. The goal of many sonification systems is the conveyance and perception of specific information related to source data as realized using sound as the principal component: an auditory display. In the same ways that computer-driven visualizations of datasets create visual analogues to the complex relationships between variables and data abstractions, sonifications represent data as audible constructs existing in time. For situations involving real-time multi-modal monitoring of continuous data streams, auditory displays working in conjunction with visual displays can provide additional attentional bandwidth for tasks ranging from the encoding of avatar motion and action in virtual space [1, 2] to correcting athletes' golf swings [3].

When sound is put to use as an expression of structured data derived from a functional process, a significant transformational and translational component is necessary to map parameters from this functional realm to parameters of sound that have the potential to express and expose specific relationships of import within the data. As audible sound itself is perceptually comprised of a great number of parameters ranging from amplitude to frequency to timbre to location in space, the organization of sounds and the organization of the parameter mappings that are associated with sound become crucial when attempting to create meaningful sonifications of complex data sets, themselves representative of complex multivariate processes. Such aesthetic and artistic exploration of data has great promise in driving new research paradigms as has been seen in recent projects begun as artistic initiatives such as Stanford University's "Brain Stethoscope" — a portable device that detects the onset of epileptic seizures through the real-time rhythmic and harmonic musical sonification of pre- and post-ictal neural activity [4].

In this light, musical sonification can be understood as the application of organizational techniques commonly attributed to musical composition and performance towards the grouping and structuring of sound generated from a data source. Musical characteristics such as pitch, rhythm, melody, harmony, timbre, tempo, spatialization, attack, decay and intensity can be mapped to raw parameters or calculated attributes of datasets, allowing for the creation of sonic output that we hear as musical in nature while still capable of conveying a great deal of information. Taken one step further, musical structures, such as the transformation of organized or generated note material, or the application of continuous changes to instrumental parameters responsible for shaping the arc of a musical phrase can also be driven by analyzed data.

Application of musical structures and techniques in this mapping process is experimental and compositional. Compositional process focuses at different levels of detail at different points in that process, ranging from low-level note-to-note attention to high-level structural attention. Similarly, sonification can range from mapping sound directly on discrete data points parsed over time to sonification of procedures in transformational algorithms. Bovermann et al. [6] further distinguish between the sonification of algorithms by juxtaposing merely inputs and outputs versus inputs, outputs, and all intra-algorithmic transformational steps between. They also propose a distinction for *operator based sonification*, where scientific models are directly embedded in sonic mapping functions.

The importance of these distinctions becomes particularly clear when comparing sonification of processed, structured data vs large sets of unstructured data. In the case of unstructured data, there may not be an inherent model with which to scaffold sonic or visual mappings and any data traversal method becomes an even more significant force in sonic signification. Significant questions around software platforms for sonification and visualization include whether the platforms are used for research-oriented exploration or post-research display of findings (presentation). Another question is that of whether the end-user is seeking aesthetic exploration (artistic) or empirical knowledge (scientific) or some mix of the two. An ideal software toolkit for data perceptualization would

https://doi.org/10.21785/icad2017.045

Figure 1: An example of visualization in *madBPM*.

allow for productive research and experiments by laboratories and artists, while also allowing that research to be presented live at professional meetings and artistic performances. The authors have produced a software toolkit and model for data perceptualization that emphasizes user-defined "behavioral abstractions" to improve sonification software flexibility and extensibility. This model has been implemented in the MadBPM software platform to create a unified research environment for both creative and analytical explorations of data through perceptualization.

2. MADBPM

The *madBPM* software platform is built around a specific model that emphasizes actions and procedures. Sonfication and visualization in this platform is realized by end-users who define code-based objects that describe data-flow and logic in the traversal of data and mapping to sound or visuals. The sound synthesis is provided by the SuperCollider sound and music programming language [5], but the modular design allows for different backends capable of Open Sound Control messaging to be used. The software is written in C++ and relies on openFrameworks[1] for the visualization functionality.

We assume that musical structure and form can be utilized to not only represent characteristics of biological processes but more importantly also to aid researchers in discovering potentially interesting and important relationships previously hidden within complex data sets. From this assumption, we designed *madBPM*, a modular sonification and visualization platform that allows for the rapid prototyping and display of sonic and visual mappings. Initially developed for a project focused on the identification of key biological data points within the process of biosynthesis for high value chemicals, *madBPM* was designed as a modular toolkit, capable of interfacing with existing audio engines, visual coding languages and customized data ingestion modules. In its current state *madBPM* is linked to the *SuperCollider* [SC] sound and music programming language [5] and the *openFrameworks* visual programming language[2].

In the following sections, we describe some of the most important architectural features of the *madBPM* software environment and data perceptualization model. These features are described in the context of the original research project from which the software environment emerged.

2.1. Data Perceptualization in *madBPM*

Auditory and visual mappings from datasets are experimentally and contextually derived. These mappings are further impacted by the initial state of the data being mapped—for instance, whether the dataset contains errors or invalid data-points. Mappings may need to account for these, or the data may need to be pre-filtered before sonification and visualization. Software environments can position one or the other approach as always necessary by design, precluding the use of unstructured data sets. *madBPM* is designed to allow for flexibility in the kinds of data sets that might be processed in the environment by emphasizing its three layers of behavioral abstractions: 1) *program-level logic*, 2) *data traver-*

[1] `http://www.openframeworks.cc`

[2] `http://openframeworks.cc`

sal/parsing, and 3) *audio/visual mapping*. In the *madBPM* software environment researchers, artists, or other users generate results by defining transformational schemes at each of these three levels of abstraction. In the last layer, audio/visual mapping, data is transformed into parameters of sonic and visual events. The second layer defines schemata for data-sets to be algorithmically traversed. In the final and "top" layer, changes in the two lower layers of abstraction can be automated. Each of these layers is described more thoroughly in section 3.

2.2. Software Environment

madBPM makes extensive use of *openFrameworks* in its architecture. Users of the platform are initially presented with a lean graphical user interface [GUI] comprised of three key components (Fig. 2). The first of the components is a pane displaying the structured data files automatically loaded at startup, described in Section 2.3. Each entry in this pane represents a "tag" that describes a subsection of the data. Clicking on tags that appear in this list selects and highlights that tag, while leaving tags that exist within that subset selectable and making tags not represented in the subset unselectable (grey, non-clickable).

The second component, at center screen, is a GUI panel that displays the console output of the managed SC process. This panel allows users to get reference information from SC or debug unexpected behavior from within the platform during any development or testing.

A third component of the platform interface features a "utility bar" like structure near the bottom of the screen. The GUI is extensible from source code and the utility bar is a potential non-intrusive spot for buttons or short-hand reference information. Currently, the bar features a color indicator box representing the connection state with SC, a Frames Per Second meter, and a button "new collection". Once a series of tags have been selected from the first GUI component (the tag list), activating the "new collection" button triggers the data from those corresponding files to be combined into a collection (ordered set) for sonification and visualization. Visualization is drawn behind the GUI, which can be hidden by a hotkey.

2.3. Structure of Data

Data is currently read by the platform from partially pre-processed CSV files. These files provide both raw data in labeled columns, but also tags that identify the relationship between each file. From the research aims of the initial phases of the project, these files communicate analyzed information from lanes on an electrophoresis gel image. The labels that identify the gel image and each lane within that image are the tags that are applied to files and subsets of data within those files.

When *madBPM* reads these files, they are organized internally into "gel_Lane" objects holding the raw data and describing their tag relationships to the platform. After tag filters have been applied and a new collection is constructed, all internally stored gel_Lanes matching the query are aggregated into a "gel_Collection" object that is used by the platform for perceptualization. "gel_Collection" objects present themselves to the platform like a multidimensional iterator that are held by a Ranger object and parsed by Walker objects (Section 3). Once gel_Lane objects are stored in memory, each gel_Lane object merely wrap immutable data and are used to create user-defined gel_Collection objects.

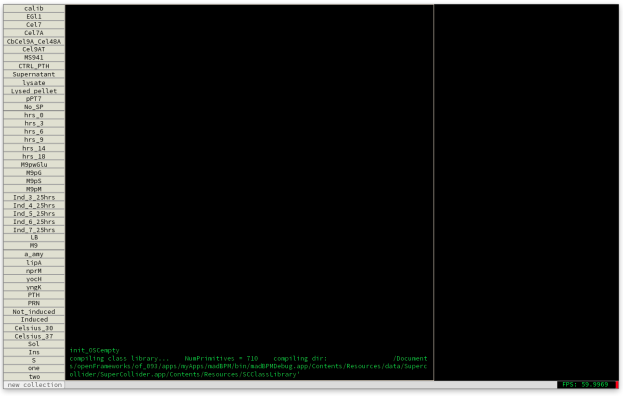


Figure 2: *madBPM* GUI at startup. At left, the tagged data filtering pane. At center is the console output of *SuperCollider*. At the bottom is the "utility bar", featuring a button to build an active collection from the selected data tags, a Frames Per Second meter, and, at far right, a color status icon indicating whether the platform has successfully initialized *SuperCollider*.

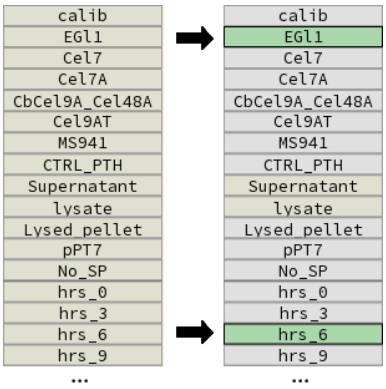


Figure 3: Detail of the tag list selection and filtering. Selecting tags filters out data that do not exist in all selected categories. Grey tags do not appear in the specified subset, while yellow tags may be selected to further constrain the set.

2.4. Sound Backend API

While the structure of visualization mappings are completely internal to the C++ application core, sonification mappings are represented within *madBPM* as objects in two ways: Sounder objects in the platform's C++ source and corresponding sibling Sounder objects in SC source. Sounder objects are encapsulations of specific algorithmic behaviors for the transformation of data into sonic or musical material. To facilitate the communication between the platform's internal representation and the SC backend representation, all data is transferred via Open Sound Control UDP messages and conforms to a strict API. Messages are directed to an address, which may be on the same computer or on another machine across a network, and are constructed as a command string followed by corresponding arguments. All arguments after the command specifier are tagged by preceding the argument with a string descriptor beginning with a ':' character.

The application level commands sent from the core platform to SC are comprised of:

- `loadSynthDefs`
  Load the platform's SC synthesis definitions.
- `create` {`:cls :id`}
  Create a new Sounder object of type `:cls` with class-specific unique `:id`.
- `updateParams` {`:cls :id :vol :rf :rv :mw`}
  Update the Sounder object of type `:cls` and `:id` with the values `:vol`, `:rf`, `:rv`, and `:mw`.
- `remove` {`:cls :id`}
  Remove the Sounder object of class `:cls` and with `:id`.
- `shutdown`
  Stop all active sounders and shutdown the process.
- `funcDefinition` {`:cls :sel`}
  Reply to *madBPM* front-end with a description of `:cls`'s function named `:sel`.

3. BEHAVIOR AND STRUCTURE IN *MADBPM* PROGRAMS

The most important design feature of the *madBPM* platform is that of *behaviors*. These are emphasized in three key levels of abstraction: program-level logic, data-parsing behavior, and data perceptualization algorithms or mappings. Each of these levels of behavior are represented within the platform as objects which describe their function over time (Fig. 5). From the beginning, *madBPM* was intended to aid in lab research data-oriented artistic inquiry, but also to enable both real-time professional presentation and artistic performance. Each of these levels of behavioral abstraction aim to address high and low level structural concerns for any of these contexts. Users define the collections, subsets of the data based on selected tags, to be sonified and visualized, and these collections are passed to the lower levels of the behavioral object hierarchy. By asking users to explicitly define the data parsing and meta structure (program-level logic), the platform is flexible enough to allow work with both un- or pre-processed datasets, structured or unstructured data, or multiple forms of data segmentation and tagging. At the current state of the project, these objects are still defined in C++ source code and compiled into the platform.

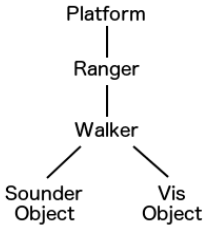


Figure 5: The behavioral object hierarchy. At the top level, the platform references a Ranger object, which defines a "program-like logic" executing over time. Ranger objects own at least one Walker object, each parsing data collections. Walkers communicate the data they parse to the Sounder objects and Vis objects they own.

3.1. Sounder Objects and Vis Objects

At the lowest level of the behavioral hierarchy are Sounder objects and Vis objects. These objects define specific mappings and algorithms for the transformation of data into visual elements on the screen or sound through speakers. These objects do not traverse data, nor do they define the rate at which data is accessed. These objects receive data from parent (Walker) objects and respond to them according to their defined behavior. All Sounder objects are polymorphic relatives of a base Sounder class, while Vis objects are similarly related to a VisObject base class. Both classes receive data from and interface with parents in identical ways: Sounders and Vis objects respond to their parent Walker object's call to an update function that accepts all relevant perceptualization data. Since Sounder and Vis objects encapsulate self-contained visualization and sonification algorithms, these objects may range from simple one-to-one mappings to much more complex real-time statistical models.

3.2. Walker Objects

Walker objects are specifically encapsulated defined behaviors for iteratively parsing gel_Collections they reference. Example behaviors might include: forward sequential traversal, visiting each gel_lane in sequence and every value in the lane; reverse sequential traversal, opposite of forward sequential; minimum to maximum traversal, visiting elements across each or all lanes from lowest value to highest; or, selective traversal, visiting every lane in the collection and updates Sounders and Vis objects only for certain values.

In the hierarchy of behavioral objects, Walkers communicate the values they visit in the data with the Sounder and Vis objects they have references to.

3.3. Ranger Objects

Ranger objects encapsulate "program-like logical structure", and interface directly with the platform and Walker objects they own. Ranger objects are analogous to the role of a musical ensemble conductor. Typically, only one Ranger class would be active at any given time, and these classes define meta-level structures and sequences during a professional or artistic presentation of the data

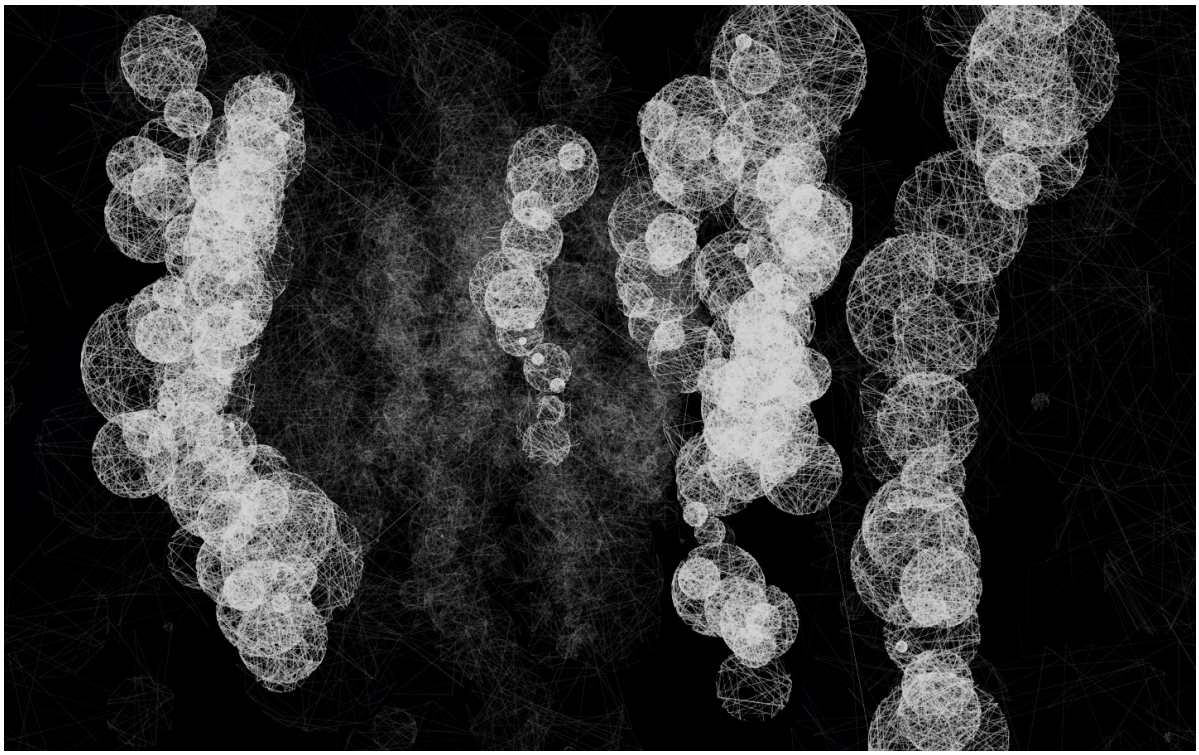


Figure 4: A closeup of the the team's "Norris" Vis_object. Data from electrophoresis gels are used to transform and extrude 3D meshes in spaces, representing density and skew as signifiers of specific trials and the distribution of their molecular weights.

perceptualization. A Ranger object, for example, might begin its operation by defining three different concurrent Walker objects and after some conditions have been met replace two of them with Walker objects of different behavior types.

These objects create the Walker objects that traverse the data, and choose which perceptualization objects those Walker objects should report to — creating, removing, or altering the relationships between these when necessary.

4. AN EXAMPLE PROGRAM

This section will describe an example hierarchy of behavioral abstractions that would define a specific operating program in the *madBPM* platform. As described in the previous section, Ranger objects encapsulate the storage and lifespan of objects that parse data streams, or the program logic of the presentation. For this example program, the Ranger object might begin operation by automatically creating two collections of data from different subsets of tags. The collections (A and B) might consist of all of the gel_Lanes tagged with {"supernatant", "M9pG", "hrs6"}, while the second consists of gel_Lanes tagged with {"Cel9AT", "lipA", "Induced"}, respectively. Next, the Ranger must associate each collection with a parsing object. Collection A could be assigned a *ForwardWalker* and collection B a *LocalMaxWalker*.

The *ForwardWalker* and *LocalMaxWalker* objects are predefined built-in Walker objects for traversing data collections they are associated with. *ForwardWalker*'s parse the each of the gel_Lanes in their associated collection in the order they are defined, and within each gel_Lane this walker visits each datum in the order defined. *LocalMaxWalker* visits each gel_Lane in its associate collection in the order defined in the collection similar to the *ForwardWalker*. However, for each gel_Lane visited, the *LocalMaxWalker* will only visit the largest local data value. The rate at which Walkers traverse the collection they are associated with is also specified by the Ranger object that defined them. The parent Ranger also specifies the action that Walker objects should take when they have reached the end of their collections. By default, Walkers that reach the end of their collection return to the start and continue parsing again. But Walkers can also be set to stop all parsing upon completion and to tell their parent Ranger that they have finished.

Instead of discontinuing parsing, the example Ranger will allow the default looping behavior and keep track of its own timing clock. Now that the program consists of subsets of data and behavioral abstractions that define how to parse them, the Ranger must associate sonification and visualization mappings for the Walkers. Perceptualization algorithms are encapsulated in Sounder and VisObjects. The example Ranger will create a *ScaleSounder* for both Walker objects, but it could provide *ForwardWalker* with a *GelBars_VisObject* (Fig. 1) and a *LocalMaxWalker* with a *NorrisMesh_VisObject* (Fig. 4). The drawn output of both visualization objects are overlaid on the same screen space. Based on each Walker object's specified sample timing, the objects poll their next data point and deliver that data to their connected mapping objects, both Sounders and VisObjects.

In this example program, the Ranger might use an internal timer to sequence changes in data collections, data parsing algorithms, and perceptualization mappings. For example, after 5 minutes have elapsed the Ranger could fade out the *ScaleSounder*

attached to the *LocalMaxWalker* and replace it with a *TimbreShapeSounder*[3]. After a few more moments, the example Ranger might remove *ForwardWalker* from the program, replacing it with a slightly altered clone of the *LocalMaxWalker*.

It is also be possible for the program flow in the Ranger to change based on conditional logic. An example mentioned above involved possibly removing a Walker once it had completely parsed a data collection instead of looping again over the data. Another possibility, however, is that Walkers that encounter data values within specific ranges could instigate structural changes in their parent Ranger. For example, if a Walker iterating over a collection encounters a value that is near a given molecular weight and has a localized intensity above a given threshold, the Ranger could respond by removing some Walkers and generating other new collections, parsing algorithms, and data mappings. The resulting change affects both the aural and visual data mappings, but also the logical structure of the analysis program. This flexibility could possibly provide a means of exploring perceptual feature optimization automatically through behavioral objects.

5. CONCLUSIONS AND FUTURE WORK

The artistic nature of musical sonification is a key element in the future plans for the *madBPM* platform, allowing our team to pursue both artistic and diagnostic goals using the project. Working in conjunction with members of the Biological Sciences department at Rensselaer Polytechnic Institute our team is investigating methods of multi-modal sonification and visualization using *madBPM* to allow researchers to better understand relationships between proteins used in the synthesis of high-value chemicals. *madBPM* allows both scientific researchers and artists to process and map data parameters from recent experiments quickly and efficiently to parameters of sound ranging from low-level synthesis techniques to higher-level organizational or compositional parameters. In this manner we envision a series of sonification and visualization experiments that analyze data sets from multiple viewpoints, allowing for fresh new looks into the data itself.

At the same time, the use of biological data as the progenitor of data-driven artworks is central to the *madBPM* project, allowing composers and visual artists to experiment with biological data in the creation of multi-modal artworks. An exhibition of such works is currently being planned at Rensselaer Polytechnic Institute's Collaborative-Research Augemented Immersive Virtual Environment[CRAIVE] Laboratory to showcase how biological data can inspire art, as well as how art can inspire research using biological data.

Within *madBPM*'s technical implementation, the project has clearly defined future milestones including:

- Implementation of a Domain Specific Language (DSL) for real-time scripting and definition of Rangers, Walkers, and Sounder/VisObjects
- Support for real-time data streams and ad-hoc data models
- Expanding the existing support for running the platform as a networked application

The creation of a DSL for real-time scripting of the *madBPM* platform will allow for more rapid research prototyping and expressive aesthetic data explorations. Currently, the team has proposed a DSL model that would focus on defining the high level behavioral abstractions, allowing users to customize data mapping and program logic at the application's runtime. The scripting language specification would require a balance between low level access to data and functions that can reach the level of composition — creating complex behaviors from underlying behavioral components.

madBPM uses a data model that is derived from electrophoresis gels in biosynthetic chemical research. In the course of implementing a data importing framework, the team found explored many useful ways of tagging, storing, and passing data around the platform. Another proposal for the future development of *madBPM* is the generalization of these methods for different types of data models and streams. In particular, the team would like to implement a pipeline for real-time data streams with data models that can be defined or redefined on the fly (in coordination with the DSL milestone).

The final proposed milestone for future development is expansion of the *madBPM*'s networking capabilities. As the size of data archives and complexity of algorithms grow, it often becomes necessary to distribute computational workloads amongst networked computer nodes. This is especially true for professional and aesthetic presentation of data perceptualization, where timing and reliability can be crucial. *madBPM*'s use of Open Sound Control already leverages a balance in speed and reliability and the possibility of communicating across a network between front and back ends. Proposed future development would break processes down into smaller units for concurrent distributed computation, allowing visual and audio display to be broken up amongst several monitors and speakers.

[3] *TimbreShapeSounder*s use filters to reshape the spectral characteristics of a sonic drone texture

6. ACKNOWLEDGEMENTS

The work described here was made possible through a grant from the Rensselaer Polytechnic Institute's Knowledge and Innovation Program (KIP).

7. REFERENCES

[1] R. Hamilton, "Musical Sonification of Avatar Physiologies, Virtual Flight and Gesture," *Lecture Notes in Computer Science: Computer Music Multidisciplinary Research (CMMR) Journal*, Springer-Verlag, Heidelberg, Germany, 2014.

[2] R. Hamilton, *Perceptually Coherent Mapping Schemata for Virtual Space and Musical Method*, Ph.D. Thesis, Stanford University, 2014.

[3] M. Kleiman-Weiner and J. Berger. 2006. "The Sound Of One Arm Swinging: A Model For Multidimensional Auditory Display Of Physical Motion." *International Conference on Auditory Display*, London, UK. http://www.mit.edu/ maxkw/pdf/kleiman2006sound.pdf

[4] G. Slack, "Hearing a Seizure's Song", *Discover Magazine*, online: http://discovermagazine.com/seizure, May, 2014.

[5] T. Bovermann, J. Rohrhuber, and A. de Campo, "Laboratory Methods for Experimental Sonification," in T. Hermann, A. Hunt, and J.G. Neuhoff, editors, *The Sonification Handbook*, Logos Publishing House, Berlin, Germany, pp. 237–272.

[6] J. McCartney, “Rethinking the Computer Music Language: SuperCollider”. *Computer Music Journal*, vol. 26, no. 4, pp. 61–68, Feb. 2002.

HUMAN AND MACHINE LISTENING OF SEISMIC DATA

Arthur Paté, Benjamin Holtzman, Felix Waldhauser, Douglas Repetto John Paisley

Lamont-Doherty Earth observatory
Columbia University
Palisades, NY, USA
pate@ldeo.columbia.edu

Department of Electrical Engineering
Data Science Institute
Columbia University
New York, NY, USA

ABSTRACT

Geothermal energy mining consists of injecting cold water into hot rocks in order to create micro-fractures allowing heat to be extracted and converted into electrical energy. This water injection can trigger several rock fracture processes. Seismologists are facing the challenge of identifying and understanding these fracture processes in order to maximize heat extraction and minimize induced seismicity. Our assumption is that each fracture process is characterized by spectro-temporal features and patterns that are not picked up by current signal processing methods used in seismology, but can be identified by the human auditory system and/or by machine learning. We present here a pluridisciplinary methodology aimed at addressing this problem, combining machine learning, auditory display and sound perception.

1. INTRODUCTION

Introduced in the 1960s for research purposes [1, 2], the transformation of seismic data into sounds has remained in use until the present day, though often for education or artistic use [3, 4, 5, 6]. However recent studies [7, 8, 9, 10, 11, 12, 13] have demonstrated the potential of auditory display in seismic research, using the power of the human auditory system to recognize patterns in signals, and to produce alternative signal descriptions that might provide seismologists with new insights and hypotheses. Concurrently, there is a recent trend in seismic research to use machine learning techniques [14, 15, 16, 17] for automatic classification purposes, bringing interesting results to the field of seismology.

Earthquakes are mainly caused by tectonic stress occurring in the Earth's crust and upper mantle. However, they can also be triggered by anthropogenic activity ("human-induced seismicity"). Human-induced earthquake may be the result of high pressure water injection into the ground, as it is found in Oklahoma during waste water disposal activity. Earthquakes can also by a by-product of human activities in geothermal fields, where water is allowed to percolate through the rock and pick up its heat, which is then used to generate electricity. Injecting under hydrostatic pressure cold water into a hot reservoir can drive different fracture mechanisms: hydraulic fracturing (microcracking driven by elevated fluid pressure), frictional sliding on existing faults (triggered by fluid pressure changes), and thermal cracking (driven by changes in volume of the rock and fluid as their temperatures change). One current challenge in the field of seismology is to be able to identify these fracture mechanisms from the seismic signals, possibly in real time, so we can have control on them in order to maximize the heat extraction while minimizing the fluid pumping and induced seismicity.

In this project we focus on the earthquakes occurring at an active geothermal reservoir at "The Geysers" in Sonoma County, CA, USA[1].

2. METHOD

The main question is the identification of fracture processes from the physical measurement (*i.e.* seismic recordings). Our assumption is that each fracture process is characterized by spectro-temporal features and patterns that are not picked up by current signal processing methods used in seismology, but can be identified by the human auditory system and/or by machine learning. We are designing an experimental method for addressing this question, bringing together psychoacoustics and machine learning. To what extent can the human auditory system identify fracture processes? Can we teach computers to identify fracture processes?

2.1. Unsupervised machine learning

The first step of the method aims to reduce the large dataset to a smaller set of signals which are typical of this dataset (*i.e.* which sample effectively the dataset), so that they can be processed by human listeners in a reasonable amount of time. Indeed, earthquakes happen constantly at the Geysers, making the 3 year-long catalog of seismic data we focus on encompass more than 46,000 events. Machine learning techniques must be used in order to reduce the dataset to its most typical elements, in reasonable number, so that they can be given to listeners.

Because nothing is assumed about the patterns that are to be found in the signals, an unsupervised approach to machine learning is used. In particular the machine is not asked to extract particular features on which the clustering would be based (in the contrary to previous studies using machine learning technique in the field of seismology, *e.g.* [14]). Because the subset extracted via the machine learning is aimed at being transformed into sounds, the spectrogram representation of the signals is preferred to time series, assuming this representation of the data can be linked more easily to how humans perceive sound than raw waveforms. The unsupervised machine learning is implemented as follows:

[1] http://esd1.lbl.gov/research/projects/-induced_seismicity/egs/geysers.html

https://doi.org/10.21785/icad2017.047

1. Learn patterns on the dataset through Non-Negative Matrix Factorization (NMF) [18, 19] on spectrograms. Each spectrogram can be reconstructed via the product of a matrix of activation coefficients (one for each spectrogram, smaller dimension than the spectrogram) and of a dictionary of patterns (a matrix common to all spectrograms);
2. Reduce the dimension of the activation matrices through Hidden Markov Models (HMM) [20];
3. Cluster the HMM-modeled activation matrices with the K-means method [20]. The machine is asked to define clusters of similar events (*i.e.* "close" vectors, according to a certain distance). Again, no hint is given to the machine on how to define these clusters. The number of clusters asked to the algorithm is defined by the user (see Sec. 3).

The only tuning that was done during the machine learning phase was the adjustment of the number of clusters. Having started arbitrarily with 10 clusters, we are investigating the effect of changing the number of clusters from 2 to 20.

The K-means algorithm assigns to each cluster a centroid. Spectrograms are then put in the cluster minimizing a distance to the centroid. This spectrogram/centroid distance can be used in order to assess the "typicality" of each spectrogram in a cluster: Spectrograms closer to the centroid are thought to be more typical of the cluster than further spectrograms. These prototypes will be selected for the listening tests, assuming that they are sufficient to provide a good and comprehensive overview of the clusters.

2.2. Audification of selected data

Seismic and sonic waves being very similar, the easiest and most conservative sonification technique is "audification" [21, 10, 11]. This technique consists in just changing the time-scale of the time series data (in our case doing a time-compression, equivalent to moving the infra-sonic frequency content of the seismic signal back up to the audible range, or to increasing the sampling frequency). Selected signals will be audified for the listening tests.

2.3. Listening tests

The interpretation of the clustering and of the criteria of "blind" machine categorization will be made through listening tests conducted with the audified data. Building on previous results that showed the potential of humans to assess audified seismic data [11, 13], the audified sounds will be the stimuli of a free sorting task. In such a task, the participants are asked to categorize sounds according to the perceived similarity. Each category then has to be verbally described, as in [13, 22, 23]. Such a categorization and verbalization task has a double purpose. First, we will check if machine and human categories agree, *i.e.* if items from the same machine-produced cluster are perceived as similar (grouped into the same category), and if items from different clusters are grouped into distinct categories. Second, the verbal description of each category gives us access to the criteria chosen by the listeners to proceed to the grouping of sounds. We will seek to interpret and link these criteria with audio descriptors (*e.g.* those defined in [24]) that can be computed on the spectrograms, but also on the spectra and waveforms, if applicable.

The abovementioned listening tests will be conducted both in a laboratory setting, and on a crowd-sourcing internet platform. Initial results from informal listenings demonstrate clear sonic differences among clusters.

2.4. Supervised machine learning

Once the signal features that are relevant to the listeners will be identified (and we assume that some of them will presumably correspond to features described in [13]), they will be computed for every signal in the dataset to check if categories produced by listeners can be retrieved by the machine. Eventually, these features — that are expected to be related to fracture processes or path effects — will feed a supervised machine learning onto another dataset: the algorithms will be fed with labelled data and will be trained to identify and categorize new data, based on criteria that are relevant from the rock mechanics point of view.

3. FIRST RESULTS

Our results from unsupervised learning show that if the number of clusters is higher than 4, the clustering comes up with one cluster gathering all events of higher magnitude. Such a high-magnitude cluster is not very informative, since the estimation of the magnitude of an event is a relatively easy task that does not need machine learning or auditory techniques. Therefore results are presented here with 2 clusters asked to the machine. These two clusters are referred to as C1 and C2 in the following.

The clusters are not characterized by any spatial criteria (they are not different faults), as Fig. 1 shows. In other words, seismic events are not clustered according to their location with respect to the source location (this would have been the case if *e.g.* one cluster had clustered all events north of the station). In a similar way, the clustering is not done according to the depth of the events (*i.e.* deep and shallow events are not separated out). However, the clustering turns out to be based on the occurrence time of the seismic events. Fig. 2 shows histograms of the occurrence times of events in each of the two clusters. This graphical representation shows distributions of dates that seem to slot very well together. This may suggest that similar physical phenomena are taking place during different periods of time. This may also be related to the history of water injection, potentially changing the fracture processes that are triggered, or the steam/liquid water balance in the rocks. We expect listening experiments to provide hints to help better understand the processes at work.

4. ACKNOWLEDGMENTS

Waveform data for this study were accessed through he Northern California Earthquake Data Center (NCEDC), doi:10.7932/NCEDC.

The authors thank Brian Bonner, Mark Warner, Kurt Nihei, Roland Gritto, Seth Saltiel, Heather Savage, Paul Johnson and Yen Joe Tan for fruitful discussion.

This project is funded through Columbia University RISE (Research Initiative in Science and Engineering) grant awarded to Ben Holtzman.

5. REFERENCES

[1] S. S. Speeth, "Seismometer sounds," Journal of the Acoustical Society of America, vol. 33(7), pp. 909–916, 1961.

[2] G. E. Frantti and L. A. Levereault, "Auditory discrimination of seismic signals from earthquakes and explosions,"

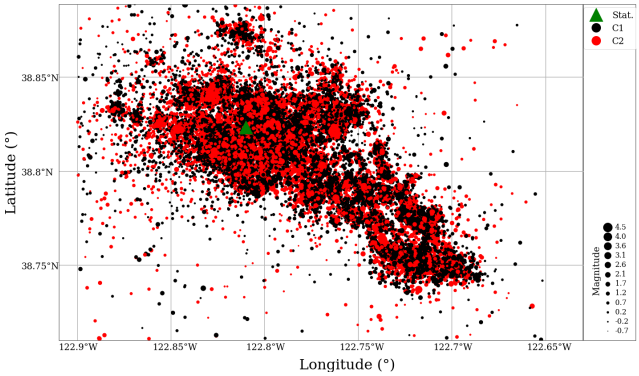


Figure 1: Map of the events (dots) in the catalog, sorted by clusters as indicated by the color code. The algorithm was asked to produce 2 clusters, named C1 (events in C1 in black color) and C2 (events in C2 in black red). The size of the dots is proportional to the magnitude of the corresponding events. The seismic station is indicated with a green triangle in the middle of the area.

Bulletin of the Seismological Society of America, vol. 55(1), pp. 1–25, 1965.

[3] B. Holtzman, J. Candler, M. Turk, and D. Peter, Seismic Sound Lab: Sights, Sounds and Perception of the Earth as an Acoustic Space. London, UK: Springer Verlag, 2014, pp. 161–174.

[4] D. Kilb, Z. Peng, D. Simpson, A. Michael, M. Fisher, and D. Rohrlick, "Listen, watch, learn: Seissound video products," Seismological research letters, vol. 83(2), pp. 281–286, 2012.

[5] M. Meier and A. Saranti, "Sonic explorations with earthquake data," in Proc. of the 14th International Conference on Auditory Display (ICAD), Paris, France, 2008.

[6] Z. Peng, C. Aiken, D. Kilb, D. R. Shelly, and B. Enescu, "Listening to the 2011 magnitude 9.0 tohoku-oki, japan, earthquake," Seismological research letters, vol. 83(2), pp. 287–293, 2012.

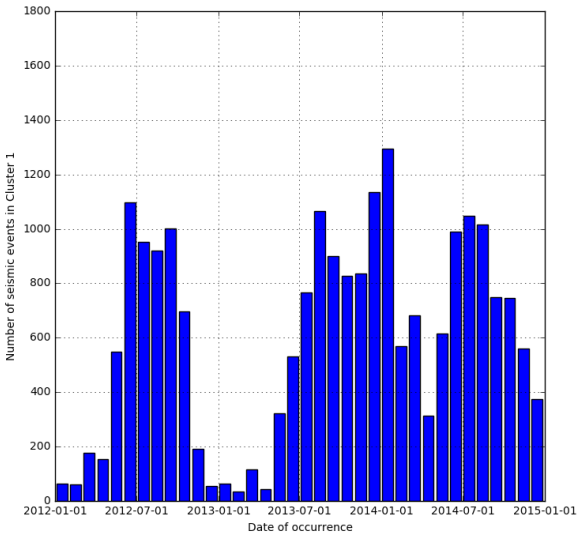
[7] C. Hayward, Listening to the Earth Sing. Boston, MA,USA: Addison-Wesley, 1994, pp. 369–404.

[8] F. Dombois, "Using audification in planetary seismology," in Proc. of the International Conference on Auditory Display (ICAD), Espoo, Finland, 2001.

[9] ——, "Auditory seismology: On free oscillations, focal mechanisms, explosions and synthetic seismograms," in Proc. of the International Conference on Auditory Display (ICAD), Kyoto, Japan, 2002.

[10] L. Boschi, A. Paté, B. Holtzman, and J.-L. L. Carrou, "Can auditory display help us categorize seismic signals?" in Proc. of the 21st International Conference on Auditory Display (ICAD), Graz, Austria, 2015, pp. 306–307.

[11] A. Paté, L. Boschi, J.-L. L. Carrou, and B. Holtzman, "Categorization of seismic sources by auditory display: A blind test," International Journal of Human-Computer Studies, vol. 85, no. 85, pp. 57–67, 2016.



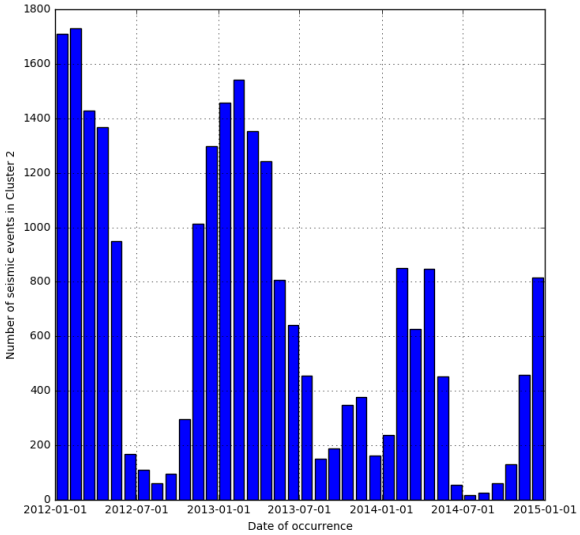


Figure 2: Histogram of the time of occurrence (grouped by months) of events in clusters C1 (top) and C2 (bottom).

[12] P. Dell'Aversana, G. Gabbriellini, and A. Amendola, "Sonification of geophysical data through timefrequency analysis: theory and applications," Geophysical Prospecting, pp. 1–12, 2016.

[13] A. Paté, L. Boschi, D. Dubois, J.-L. L. Carrou, and B. Holtzman, "Auditory display of seismic data: On the use of experts' categorizations and verbal descriptions as heuristics for geoscience," Journal of the Acoustical Society of America, vol. Accepted on Feb. 15th, January 2017.

[14] F. Provost, C. Hibert, and J.-P. Malet, "Automatic classification of endogenous landslide seismicity using the random forest supervised classifier," Geophysical Research Letters, vol. 44, pp. 113–120, 2017.

[15] G. Curilem, J. Vergara, G. Fuentealba, G. A. na, and M. Chacòn, "Classification of seismic signals at villarrica volcano (chile) using neural networks and genetic algorithms," J. Volcanol. Geotherm. Res., vol. 180(1), pp. 1–8, 2009.

[16] F. Dammeier, J. R. Moore, C. Hammer, F. Haslinger, and S. Loew, "Automatic detection of alpine rockslides in continuous seismic data using hidden markov models," J. Geophys. Res. Earth Surf., vol. 121, pp. 351–371, 2016.

[17] P. B. Quang, P. Gaillard, Y. Cano, and M. Ulzibat, "Detection and classification of seismic events with progressive multichannel correlation and hidden markov models," Comput. Geosci., vol. 83, pp. 110–119, 2015.

[18] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, vol. 401, pp. 788–791, 1999.

[19] ——, "Algorithms for non-negative matrix factorization," in Proc. of the 13th International Conference on Neural Information Processing Systems, Denver, CO, USA, 2000, pp. 535–541.

[20] C. M. Bishop, Pattern recognition and machine learning, ser. Information Science and Statistics. New York, NY, USA: Springer Verlag, 2006.

[21] F. Dombois and G. Eckel, Audification. Berlin, Germany: Logos Verlag, 2011, ch. 12, pp. 301–324.

[22] A. Paté, J.-L. Le Carrou, B. Navarret, D. Dubois, and B. Fabre, "Influence of the electric guitar's fingerboard wood on guitarists' perception," Acta acustica united with acustica, vol. 101, pp. 347–359, 2015.

[23] C. Guastavino, "Categorisation of environmental sounds," Canadian journal of experimental psychology, vol. 61, pp. 54–63, 2007.

[24] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," Journal of the Acoustical Society of America, vol. 130(5), p. 2011, 2902–2016.

Paper Session 5
Games

USING AUDITORY DISPLAY TECHNIQUES TO ENHANCE DECISION MAKING AND PERCEIVE CHANGING ENVIRONMENTAL DATA WITHIN A 3D VIRTUAL GAME ENVIRONMENT

James Broderick, Dr Jim Duggan, Dr Sam Redfern

College of Engineering and Informatics,
National University of Ireland, Galway,
University Road, Galway City, Galway, Ireland
j.broderick4@nuigalway.ie

ABSTRACT

When it comes to understanding our environment, we use all our senses. Within the study and implementation of virtual environments and systems, huge advancements in the quality of visuals and graphics have been made, but when it comes to the audio in our environment, many people have been content with very basic sound information. Video games have strived towards powerful sound design, both for player immersion and information perception. Research exists showing how we can use audio sources and waypoints to navigate environments, and how we can perceive information from audio in our surroundings. This research explores using sonification of changing environmental data and environmental objects to improve user's perception of virtual spaces and navigation within simulated environments, with case studies looking at training and for remote operation of unmanned vehicles. This would also expand into how general awareness and perception of dynamic 3D environments can be improved. Our research is done using the Unity3D game engine to create a virtual environment, within which users navigate around water currents represented both visually and through sonification of their information using Csound, a C based programming language for sound and music creation.

1. RELATED WORK

Before this project, work was done using the Unity 3D game engine to create a virtual environment for collaboration and visualization of marine environmental data [1]. The system created a replication of Galway Bay and visualized surface current data across the bay, displaying the direction and speed of currents across the bay each hour over a month-long period. The goal was to show how such a system would allow for the data to be easily perceived and understood by a variety of users, who would then be able to use the system for collaborative discussion and decision making. The environment is created from LIDAR data of the seabed, creating quite a realistic representation. The ocean current is then mapped onto its correct geographical location within the environment, allowing for more data and in greater varieties than simple surface currents to be added to the system over time. The Galway Bay Model and some theories and ideas have been carried over to this project.

There are some examples of sonification of environmental data, with focus on user perception of the data rather than navigation. In DoppelLab [2], a virtual recreation of MIT's Media Lab was created in Unity. The goal of this recreation was for simultaneous representation of

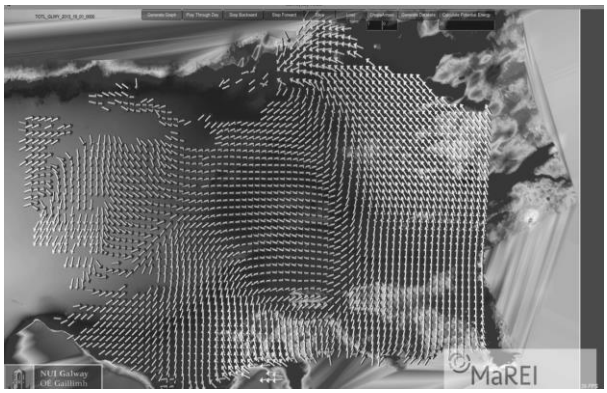


Figure 1: Surface currents visualized in previous work

environmental data captured by sensors within the building as visual and audio sources. With several hundred sensors around the building measuring humidity, temperature, noises levels, and actual captures of sound, it can be difficult to discern between different sources when relying purely on visual representations of data. The writer explored the usage of sonification to represent data as sound in addition to the visualizations. Being able to experience data both audibly and visually makes it easier for users to not only understand the environment, but easily discern where different sources of information end and begin, and how exactly they are located compared to each other. As users analyzed the virtual environment, they could better map out how busy different areas were at what times, how it affects temperature, etc. Seeing this, it is a natural to imagine that a same user could move from simply using this technique to understand the environmental data towards using this information to adjust their real-time actions or greater goal in the environment.

In fact, it has already been shown that, just like our real-life experiences, adding audio cues to our virtual environments aids our navigational ability. There have been several studies of using auditory navigation waypoints at specific goals or locations to aid users in being able to move through the environment [3][4]. Grohn and Lokki looked at measuring improvements in users finding objects within the environment using visual, audio, and audio visual cues to represent the target goals. While audio on its own was the least successful method of locating goal objects, combining visual and audio lead to users finding many more objects within a set time constraint. It was also found that users would use audio cues first to roughly locate an object before using visual stimuli for the final approach. Walker and Lindsay also looked at using auditory waypoints for navigation of an environment, specifically how users would

https://doi.org/10.21785/icad2017.012

be able to follow a path of waypoints with primarily auditory stimuli. It was found that users were well able to find and follow these waypoints, with the only common issue being users overshooting waypoints upon getting too close to them.

Both in terms of navigation and on a more complex scale, we are looking at aiding user decision making. Beyond examples of navigational decision making using auditory signals, there is also evidence of sonification aiding in other types of more complex decision making. Whether it be monitoring network traffic [5] or keeping track of multiple task prioritization [6][7], auditory display greatly enhances the user experience as a background process. Rather than distracting from the task at hand, users are simply listening for changes in their auditory background to indicate their attention is needed elsewhere. In situations where a purely visual display would hinder task completion and decision making, users can instead focus on the task at hand. Users were then not only able to handle their tasks at the best of their ability, but were better able to react to changing situations within multiple tasks simultaneously.

By better understanding user decision making, we can better create systems that aid users, especially when using new techniques for information perception. From studying instances of various navigational, task prioritization and task monitoring as seen earlier, it is understandable how exactly to fit in audio sources and measure their effectiveness in decision making. Beyond that, is important to understand how exactly users make decisions, especially as we hope to move on to more complex decision making experiments than simply navigational control. Some studies of decision making we looked at were the Beer Game, specifically with newer computerized versions of the game [8], and the Fish Banks Exercise [9]. While the Beer Game isn't as directly applicable to the modern world, it is still an interesting exercise and learning tool, especially for studying user's initial reactions and showing how small decisions cause greater effects on the rest of the system and users. It also allows students to see the importance of information gathering/sharing. The Fish Banks Exercise allows for similar studies of how poor short term decision making can damage the entire system. By using these studies and their research as to why users make the decisions they do, we hope to be able to carry over some of this knowledge to our own studies.

2. HYPOTHESIS

It's been shown by previously mentioned work that using static auditory beacons or waypoints, users are better able to navigate virtual environments. Users are also able to understand some amount of complex information represented as audio. The theory of this study is that by sonifying changing and mobile environmental data, we can aid in user navigation of these virtual environments. Rather than focusing on using auditory display for finding a user's destination, we want to look at how sonification of the potentially hazardous environment around the user can lead to better awareness of a user's location within an environment and their avoidance of hazards.

We use a reliance on hearing to keep us aware of how our environment outside of our sight changes, and this has been used in video games for a long time. Whether it be enemy footsteps, the sound of a power up activating, or even audio pings when an enemy draws close, games use auditory cues to keep users aware of their changing environment without cluttering the screen with a huge amount of visual information. Sound design such as this and the techniques that game developers use to portray this environmental information can be carried over to other virtual environments to better enhance user experience.

2.1 Example Use Case

A use case being examined is that of navigating a remotely operated vehicle (ROV) in a marine environment. Other researchers in the MaREI research group in the University of Limerick have worked on development of an ROV for marine missions, and as part were using a virtual control system for the ROV. As part of further research, they were moving this control system into a Unity 3D virtual environment and control system. When thinking of collaborative work between this and the previously mentioned Galway Bay virtual environment, we thought of how the control of the ROV could be enhanced with auditory display of parts of the environment. Visual aspects of the environment were already being displayed, but the presence of audio components could lead to users having a better perception of this remote, 3D environment.

As a first case study, this experiment will look at having users navigate a virtual rover through a marine environment. While the environment itself is a simple seabed based on LIDAR data of Galway Bay, water currents will provide the primary obstacle to navigation. These currents can change in intensity and direction, and can be in any location around the user. The aim is to sonify basic parameters of these currents so that the user has better awareness of their environment in all directions at all times, being able to tell the distance from, direction of, and strength of the currents around their position. This should lead to an improvement in navigation in timing or accuracy, as they will be less likely to stumble into a current outside of their field of view, or be surprised by a sudden change in direction or intensity. Since these currents can appear in any direction around them, sound should give better location awareness than just visual representation.

3. PLANNED SYSTEM

3.1 Unity3D Game Engine

Game engines have potential for use in non-entertainment projects for a variety of reasons. Modern game engines are built to be highly modular; where older engines were aimed at a specific type of game, modern engines are expected to be able to create anything from an independent 2D simulation game to a commercial first person shooter. They have basic components for cameras, 3D models, networking, controls, physics, etc., as well as having a huge amount of customization through creation of user scripts. All of this is aimed to make the engine applicable for a wider variety of games, but it also lets these tools be used for creation of non-gaming projects, such as the collaborative visualization tool

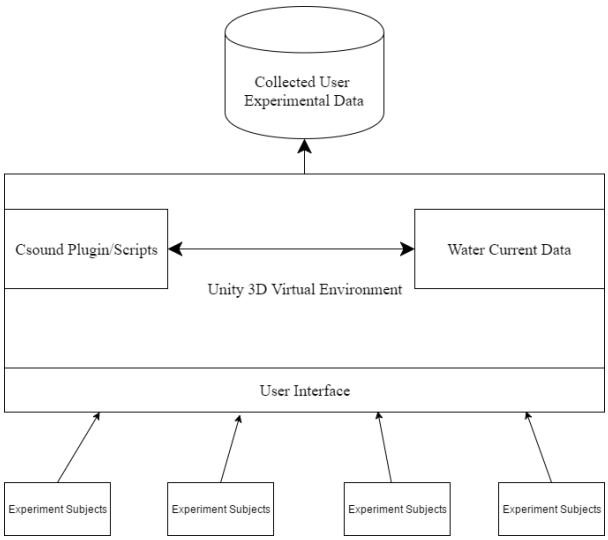


Figure 2: Architectural Diagram of planned system.

in this paper. By using a game engine for the basic setup of a system, more focus can be put on the actual functionality being created, rather than spending time recreating a rendering or physics engine. As well as this, with the growing popularity of independent games, many people are actively making tutorials and discussing game projects online, creating a wealth of information and experience.

The Unity3D game engine is the engine of choice for this research project, the reasons for which will be covered now. Most game engines provide a similar baseline functionality, with systems for rendering/visualization, sound, physics and scripting for functionality. Unity has a lot of flexibility with its scripting, as it allows for use of Unity versions of C#, JavaScript and Boo. Scripts of different languages can be used in the same project and on the same game objects with little conflict. This means there is a smaller learning curve and gives more room for coders to work in their preferred language. Unity is also highly portable, easily able to deploy on platforms such as PC, OS X, Linux, Mobile, Web, and a variety of game consoles. This portability means programs can be built to work on a variety of systems so that the maximum number of users can be supported. Unity has a huge user base, with people constantly writing tutorials, creating assets and answering questions. Additionally, projects and companies have used Unity for "Serious Games" in the past, including virtual environments [10], urban planning [11] and disaster simulation [12].

3.2 Csound

Csound is used for the auditory display requirements of the project. Csound is "a sound and music computing system" developed in 1985 at MIT Media Lab. It's a flexible way for creating computer driven music and sounds, running on a multitude of platforms. The primary reason for the usage of this sonification method is the existence of CsoundUnity, a C# wrapper for Csound to be integrated with Unity3D. Developed by Rory Walsh [14], it allows for use of Csound based instruments and sounds within Unity environments, drawing information from the Unity environment for use by Csound and extending Unity's audio API. Being able to have the modularity and flexibility of Csound directly available in Unity3D allows for a strong blend of visual and audio projects. While the sonification aspect of this first use case is relatively simple, having access to the strength of Csound means that similar projects can have a huge level of granularity when it comes to control of audio sources within their virtual environments. Hopefully by having free and easy access to a powerful 3D game engine that works directly with this software will allow for greater research into auditory display in virtual environments by new researchers.

4. EXPERIMENTAL PROCESS

The upcoming experimental process will involve users navigating an ROV through a series of currents in a Unity based virtual recreation of Galway Bay. Using mouse and keyboard controls like many generic video games, they will move the ROV through set checkpoints towards a destination. The possibility exists to also have the currents hinder or change the movement of the ROV, leading them to cause time issues as well as points penalties, but opening the possibility for smart use of the currents to save time. This gives user decision a greater depth of possibility, with more choices than simply what direction to go. While the users navigate the environment, the system can track user location data and movements, the time it takes them to reach objectives, and how often they stray into water currents, as well as the strength and direction of those currents. These factors can later be analyzed and compared.

Participants will be gathered primarily from within NUI Galway's Computer Science undergraduate classes. One known issue with auditory display is the additional learning element required. Users must be aware of what different audio means to gather useful information from it. By having mostly younger, computer-orientated participants, it is hoped that many of them will be used to basic gaming controls and functionality. This means that any learning within the environment can be focused on the auditory display side of the project. By reducing the amount of different simultaneous learning components, we can hopefully gather a truer reading of the usefulness of the sonification aspect.

As mentioned earlier, Unity's portability means the test environment can be deployed through the web. This planned experiment is aiming to use a web deployment of the system that users can connect to. Users access the test system through web browsers, log in with given details, and can perform the experiment entirely through their own machine with almost no setup. This means in addition to set computer

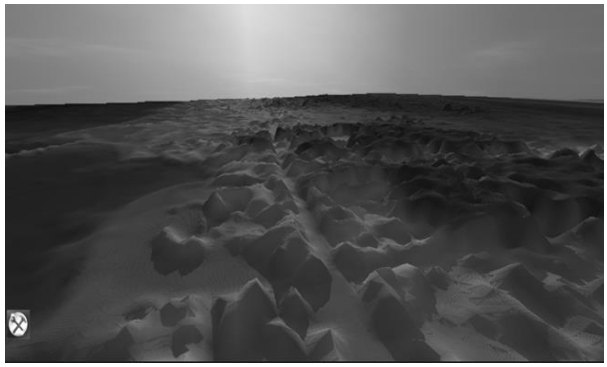


Figure 3: Galway Bay Seabed in Unity.

lab settings, more data can be garnered through remote test groups who need only a reasonable computer setup, headphones, and an internet connection. User results are stored on the server for centralized access.

The ocean current data being used in the experiment is based off of surface currents of Galway Bay, expanded into a 3D grid. The original set of data is purely surface currents, containing location information, timestamps, and information on the water's direction, strength, etc. at that position. This data is converted into a grid across the bay by converting the real world latitude and longitude coordinates into coordinates within Unity space. This means that the currents are correctly positioned in their real world equivalent location within the virtual representation of Galway Bay. For the purposes of this experiment, this data is used as a basis for rough values due to it being 2 dimensional and having a distance of approximately 1km between data points. The currents used in the virtual environment are exaggerated and more changeable than their real world counterpart to make the task more obvious for the user. However, with future work having access to more in-depth datasets and of a wider variety than ocean currents, the theory behind this experiment could be used for a wider set of case studies and ideas such as the ones listed previously.

The current information is sonified using three main parameters: Distance from the user, strength of the current, and direction of current. Using Unity's 3D sound placement, game objects in the virtual environment representing the currents will output audio sonifying the current it represents. The direction the sound appears to be coming from will adjust around the user, so users will hear currents from literally any direction around them. In addition to this, the volume of the sound grows as users get closer to the source of the audio. This means users will be able to tell what direction the current is, and roughly how close they are to it. After moving a certain distance from a sound source, audio is cut off to stop overwhelming of users with audio. The frequency of the audio changes depending on the speed of the current, so users will be able to tell how powerful a current is depending on how low or high the sound is. The combination of parameters should allow users to not only have a good idea of their full nearby surroundings, but also what directions and sounds should be prioritized. A more complicated parameter is sonifying the direction of the currents. As of now, this is using a moving sound source within the visual representation of the current. A simple Csound function is used to control the various parameters of the sound, with information on the current speed and distance from user handled by the Unity game object and fed into Csound as usable data.

One drawback of Unity's default audio is that its ability for audio spatialisation is limited. Currently it relies mostly on panning audio, which causes issues when audio sources are directly in front of or behind a user. These audio sources will sound identical, meaning users are forced to rotate their viewpoint to better orientate themselves with these sounds. This can be solved with implementation of Head Related Transfer Function (HRFT), which works to emulate how sounds are affected by our head and ears as we hear them. Part of our experimentation is to compare test groups using either form of audio spatialisation to see how great an effect it has on the user's perception of audio sources while multi-tasking within a virtual environment.

The system will have two main user types: Users who will move through the system with purely visual cues, and users who will receive a mixed audio-visual environment. There is no purely auditory level set as the goal of this research is to enhance visual systems with useful auditory displays. The first level of the system will have users trained in the system with only visual representation of currents in a simplified level. This allows users to focus on understanding the goals of the experiment and get used to the control of the ROV. The second stage will be a more complex current set up, and will also be the introduction of auditory representation of the nearby currents. This second stage is aimed at acclimatizing users to the sonification of the currents, and is primarily an additional training step. The third and final step is the primary source of results. Some users must navigate a level with only visual representations of the currents, and other users will have both visual representation and auditory representation. Performance in this step will be compared between user groups to find any improvement in navigation.

5. PLANNED WORK

The first stage of the experimental system is nearing completion. Once completed, experiments will be conducted with test groups to gather our first set of results. First will be examining changes to the current case study. Once user feedback and results have been gathered and analyzed, the method of sonifying can be further studied and improved. Different types of sound parameters may be used for sonifying environmental data in further experiments, and the specific values and distances can be fine-tuned. By using Csound and Unity, changing how the sound is represented, or how the environment handles the audio, are simple changes and can lead to more variety of experiments in the future.

There are also plans to implement support for Oculus Rift in the experiment to see if the addition of virtual reality and head tracking would have an effect on users' perception of the environment and ability to locate audio sources within the environment, as well as providing more in-depth information on how users move and react to changing sounds outside of their visual viewpoint. This would also make it easier for users to handle the back-front issue present when using simpler forms of audio spatialisation. As the experiment is built using Unity, the addition of Oculus Rift is relatively simple, with the main limiting factor being the additional equipment leading to smaller and slower testing of individuals.

There is also room for new types of experiments. Some plans are to look at combining previous work with auditory waypoints with the current system for environmental sonification. This can include measuring how additional sound objects affect user experience and understanding of the environment. As well as this, there is potential for having multiple types of environmental sonification that can be cycled through by the user in a single environment. Being able to switch between sonification of auditory hazards and audio waypoints showing directions to travel could have interesting results.

6. ACKNOWLEDGMENT

We would like to acknowledge the support of the Science Foundation of Ireland, the MaREI project and the College of Engineering and Informatics at NUI Galway.

7. REFERENCES

[1] J. Broderick, J. Duggan, S. Redfern, "Using Game Engines for Marine Visualisation and Collaboration", in *Proc of the 2016 Int. Conf. on Image, Vision and Computing (ICIVC), Portsmouth, UK, 2016, pp. 96-101*.

[2] N. Joliat, "DoppelLab: Spatialized Data Sonification in a 3D Virtual Environment", Master's Thesis, Massachusetts Institute of Technology, February 2013, retrieved from http://hdl.handle.net/1721.1/85427

[3] B. Donmez., M. Cummings, & H. Graham, "Auditory decision aiding in supervisory control of multiple unmanned aerial vehicles", Human Factors, Vol 51, Issue 5, pp 718–729, 2009

[4] D. Brock, J. Stroup, J. Ballas, "Effects of 3D Auditory Display on Dual Task Performance in a Simulated Multiscreen Watchstation Environment", Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 46, Issue 17, pp 1570–1573, 2002.

[5] P. Kaminsky, D. Simchi-Levi, "A new computerized beer game: A tool for teaching the value of integrated supply chain management. Supply Chain and Technology Management", pp, 216–225, 1998

[6] J. Whelan, "Building the fish banks model and renewable resource depletion", pp. 70, October 2001.

[7] S. Wang, Z. Mao, C. Zeng, H. Gong, S. Li, B. Chen, "A new method of virtual reality based on Unity3D", 18th International Conference on Geoinformatics, 2010, pp.1–5.

[8] A. Indraprastha, M. Shinozaki, "The investigation on using Unity3D game engine in urban design study", ITB Journal of ICT, vol. 3, issue 1, pp.1–18.

[9] S. Sharma, S. Jerripothula, S. Mackey, O. Soumare, "Immersive virtual reality environment of a subway evacuation on a cloud for disaster preparedness and response training", Proc. SPIE 9392, The Engineering Reality of Virtual Reality 2015, 939208

NAVIGATION IN AN AUDIO-ONLY FIRST PERSON ADVENTURE GAME

Adrian Jäger

Department of Media Production
OWL University of Applied Sciences
Liebigstrae 87,
32657 Lemgo, Germany
jaegerva@gmail.com

Aristotelis Hadjakos

Center of Music and Film Informatics
Detmold University of Music
OWL University of Applied Sciences
Hornsche Str. 44, 32756 Detmold, Germany
hadjakos@hfm-detmold.de

ABSTRACT

Navigation in audio-only first person adventure games is challenging since the user has to rely exclusively on his or her sense of hearing to localize game objects and navigate in the virtual world. In this paper we report on observations that we made during the iterative design process for such a game and the results of the final evaluation. In particular we argue to provide a sufficient number of unique sound sources since players do not use a mental map of the virtual place for navigating but instead move from sound source to sound source in a more linear fashion.

1. INTRODUCTION

Since the invention of the first computer games about 50 years ago, significant progress has been made. Numerous milestones were achieved from the early text-based games and sprite graphics, to the simulation of three-dimensional worlds, massively multiplayer online games and the development of virtual reality. Also for sound and music, a game developer has many options today, including recorded music and sound samples, synthesized sounds, generated music and spatial audio. Sound and music are used in almost all games today, e.g., to create a desired emotional response with sound effects and background music or to develop the story with narration. However, the human-machine interaction in mainstream games is mainly visual: The user interacts with a virtual world that is mainly graphically represented and performs actions on visual elements. While the user's actions relate to vision, the feedback is often both visual and auditive. This makes it much more difficult but not impossible to play a game without visual feedback (as we know from visually impaired people who play mainstream games with some success [1]).

In audio games, the interaction is mainly auditive. The user interacts with an auditive virtual world without (or with very little) visual feedback. Audio games have a unique aesthetic quality and can stimulate the player's imagination. Thus a good audio game should normally be equally interesting for visually impaired and unimpaired people. Audio games are of course a niche genre.

A particular challenge for audio-only first person adventure games is navigation in the virtual world. Since there are no visual cues, the player has to rely on his or her hearing skills to localize sounding objects and navigate in the world. While there are some studies on navigating in virtual environments (see Section 2.2), this paper examines navigation in the context of audio-only gaming. We report our observations that we gained during the iterative design of the game (see Section 5). We then discuss the results that we obtained in an evaluation of the game (see Sections 6 and 7).

2. RELATED WORK

2.1. Audio games

Wolf categorizes graphical computer games by genre [2]. He distinguishes the games based on the primary interaction characteristics offered to the player, the goals to be achieved, the type of control, and the player's avatar. On the basis of these criteria he describes 42 different genres, some of which also apply to audio games.

2.1.1. Text Adventures

In text adventures, the narration of the story is done by a computer voice or a recorded human voice without visual feedback. The user can select an action from a list of options. By doing this, a branch in the course of the story is chosen, which can have short- or long-term effect. An example of an audio text adventure is "Descent into Madness."[1]

2.1.2. Grid games

While many games are played in the first person perspective, some games limit the freedom of movement of the player to a fixed grid consisting of squares of unit length while the game world can still be explored relatively freely. The grid functions as a support and makes sure that game-relevant objects can not be so easily missed.

An example is "3D Snake,"[2] an adaptation of the famous "Snake" game. Using the arrow keys, the player navigates over a two-dimensional plane. When approaching the limits of the playing ground, this is announced by a characteristic sound that gets louder and louder. The fruits to be collected are recognizable by their own sound, which (again) gets louder as the distance between the player and the fruit decreases.

https://doi.org/10.21785/icad2017.033

[1] Descent into Madness: http://www.cs.unc.edu/Research/assist/et/2005/SoundsLikeFun.html, last access: Feb. 2017.

[2] 3D Snake: http://usagamesinteractive.com/freeware.php, last access: Feb. 2017.

2.1.3. First-person adventures

First-person adventures offer greater freedom than grid games: The player can move freely in the virtual world independently of a grid. Neither the length of the step, nor the angles of a rotation to the right or left are perceived as being discrete. The game principle calls for a realistic, auditive virtual reality.

The game "Slender: Lost Vision"[3] can serve as a suitable example. As the name implies, this is an audio version of the game "Slender - The Eight Pages".[4] The main character is a girl, who walks into a nightly forest out of curiosity. The girl has to collect eight pieces of paper attached to different trees without being caught by the Slenderman. The paper notes are rustling in the wind and can be heard by the player. With the right mouse button, an audio compass can be triggered that generates a sound located in the northern direction to make the orientation somewhat easier. But also the birds, which sit in the trees, help with the orientation. The game creates a lot of excitement through a buzzing, pulsating background music and sudden, frightening, loud noises that indicate the presence of the Slenderman.

2.1.4. Side scrollers

Similar to their visual counterparts, audio side scroller games run one-dimensionally from left to right. Obstacles that need to be skipped or repelled, and any auxiliary items that can be collected are announced by audio signals. These signals are usually learned beforehand in a short tutorial.

"Adventure at C"[5] works this way.The player is a programmer whose hard drive is infected with viruses. To solve this problem, he goes on a journey through his hard drive and encounters viruses, which he has to fight, holes that he must skip and other obstacles. A common challenge in audio side scrollers is limited or missing feedback, which can make it hard to learn from past runs and perform the right corrective action.

2.1.5. Rhythm and skill games

Most of the categories listed above have in common that they either follow a story or evoke a space, a world. This is different in this category: the games require a mostly rhythmic interaction in which the player has to press a button or perform a specific movement. In some cases, the timing of the interaction is not relevant but the order of the pressed buttons, wherein the time intervals therebetween may be arbitrary.

An example of such a game is "Simon".[6] The game consists of a device with four buttons, which produce different sounds. The built-in computer first plays a series of few sounds, which must then be reproduced by the player on the keys. If successful, the melody is extended by one note at the end and the player must play the now expanded melody from beginning to end. This scheme continues until the sound series is complete or the player makes a mistake.

2.1.6. Board games

Many board games have been adapted for the visually impaired. In those games, all relevant information is announced by the computer, e.g., the name of the player who is to move next, the card placed, the cards owned by the player or the diced numbers. Such adaptations also make it easier for impaired and unimpaired people to play board games together.

An example of an audio board games is "Blind Gamers Hearts".[7] It is an adaptation of the card game "Hearts" where a computer voice announces the active player and his placed card. The player then uses the arrow keys to scroll through his cards, each card being read aloud. The desired card is placed with the Enter key and the game continues.

2.2. Navigation in virtual audio environments

Lokki et al. [3] studied auditory navigation in virtual acoustic environments without visual feedback. They set up an experiment, where the users would navigate to multiple sound sources placed in a virtual 2D world with keyboard commands to walk forward, backwards and to turn left or right. They examined the effect of different factors, including:

- Panning methods: interaural time difference (ITD), ITD + interaural level difference (ILD) and ITD + head-related transfer function (HRTF)
- Simulated acoustic environments: direct sound only, direct sound + reflections and direct sound + reflections + artificial reverb

Lokki et al. counted how often the users found the target area and measured the time spent in the navigation task. In their experiments, the users performed best with ITD + ILD panning. The best acoustic environment for the navigation task was direct sound only, while direct sound + reflections + artificial reverb performed worst [3]. Since we wanted our game to sound aesthetically pleasing and since the differences measured by Lokki et al. are only moderate, we chose to use reflections and artificial reverberation nevertheless. Also the navigation task in the study by Lokki et al. was limited to targets located within a single room so that the reflections did not convey information relevant to the navigation task. In a more complex (and for a computer game more typical) setup with multiple connected rooms and spaces, reflections could be helpful since the user could perceive the spatial layout, e.g., by perceiving a lack of reflections from an area where a passage leads to another room. In a further study Gröhn, Lokki and Takala examined auditory and audio-visual navigation in a 3D space, where the user could move freely in all six degrees of freedom [4]. Interestingly, the users improved when they received auditive cues in addition to visual cues.

Picinali et al. created a system that lets visually impaired people explore buildings using virtual acoustics so that they get along better when they visit the building in real life [5]. The users wore headphones with an orientation sensor, which was used in combination with a joystick to navigate. Virtual sound sources were placed in the room, e.g., a group of men talking. Furthermore, the user could generate sounds (finger snapping, footsteps) to explore the sound properties of the virtual room. Picinali et al. showed that their users were able to create relatively accurate mental maps

[3] Slender – Lost Vision: http://dragonapps.org/audiogames/slenderlostvision/, last access: Feb. 2017.

[4] Slender – The Eight Pages: https://de.wikipedia.org/wiki/Slender_Man, last access: Feb. 2017.

[5] Adventure at C: http://www.vgstorm.com/aac/, last access: Feb. 2017.

[6] Simon: https://en.wikipedia.org/wiki/Simon_(game), last access: Feb. 2017.

[7] Blind Gamers Hearts: http://www.omninet.net.au/~irhumph/bghearts.htm, last access: Feb. 2017.

of the building [5]. Our users (with unimpaired sight) were not able to construct such accurate mental maps. This is consistent with a study that compared sighted and visually impaired people in a comparable audio navigation task [6]. Of course, other factors such as the differences in interaction (orientation sensor and joystick vs. mouse and keyboard) or differences in the spatial sound system could also have contributed to the different outcome.

But audio navigation in an audio-only computer game does not have to rely on real-world acoustics. Completely new mappings could be used, e.g., those proposed and examined by Lorenz and Schuster [7]. They modulated various parameters of a synthesizer (including pitch, tempo, LFO, reverberation, panning and timbre) by the distance between the mouse cursor and the target. They measured how long it took for the users to acquire the target and the distance that the mouse cursor traveled. They also computed the directional accuracy, which is the mean deviation of the angle between the optimal and the real direction of mouse movement. Based on these measurements, Lorenz and Schuster conclude that tempo, pitch and timbre performs best while "stereo panning [...] seem[s] hardly suitable for navigation tasks" [7]. Of course their navigation task was clearly different than ours. Their sonification techniques could however be applied for navigation in first person adventure games. Of course, in that case the game would not sound realistic any more.

3. SIMULATION OF ROOM ACOUSTICS

Our game is a first-person adventure. For such games we deem the quality of spatial sound reproduction to be crucial. This is even more important than in traditional video games, since the sound is the only modality. Sounds of a virtual world can be converted synthetically into binaural signals [8]. Such approaches depend, on the one hand, on the physical behavior of the sound waves, and on the other hand, on the biological presuppositions and functions of human hearing. An in-depth introduction to physical and physiological foundations can be found in [9, 10, 11, 12].

In 1999, Mueller and Ullmann conducted experiments in which reflections were simulated in a three-dimensional space [13]. The simulation was based on the idea of ray tracing, a common technique in 3D rendering. Beams emanating from a sound source are emitted. When a collision with an object such as a wall occurs, a further sound source is produced, which in turn emits further beams. This operation is performed until the specified maximum number of beams is reached. Beams that collide with the virtual representation of the listener are used to calculate the final sound. In addition, the angle of incidence or the exact location of the collision determines how the HRTF filter is calculated for the respective beam. The computers used at that time required about 13 hours for the calculation of a ten-second sample with a limitation to 10,000 virtual sound sources [13]. Today, ray tracing simulation of room acoustics is possible in real-time, e.g., by making use of the GPU [14]. Alternatively, the propagation of sound waves can be simulated. This is physically more correct but computationally more expensive. Again the GPU can be used for an efficient implementation [15].

The most common, commercially available programming frameworks for spatial audio are Aureal A3D, Creative EAX, FMOD, Microsoft DirectSound 3D, and OpenAL. These can be divided into two groups:

- The frameworks in the first group simulate a space without paying attention to the exact geometry of the space to which the simulation is applied by using prefabricated presets, e.g., with fixed impulse responses. Therefore, the calculation of the audio signals does not react dynamically to the movement of the listener in the virtual space. Such frameworks include DirectSound 3D, Creative EAX and OpenAL.
- The frameworks in the other group calculate sound reflections dynamically in real time and can therefore produce sounds that depend on the position of the listener in the virtual space. This group is mainly represented by Aureal A3D [16, p. 49] and dearVR [17]. Other systems have not been commercialized or have not achieved a certain level of commercial use. Since Aureal A3D has been discontinued, we chose dearVR as spatial audio framework for our audio game.

dearVR uses the ray tracing to determine early reflections. This is based on the actual geometry of the room in which the sound event takes place. dearVR considers only those reflections, which reach the listener after a single contact with a wall or any other object. The direct sound and the early reflections are then spatialized using a HRTF. The phase cancellations resulting from the delay in the summation of the signals are similar to those occurring in reality enough to give the player an impression of an actual three-dimensional space.

All further reflections are not calculated in this way but are produced with a reverberation effect. This reverb can be selected from a list of prefabricated presets. This reverb is calculated independently of the position and orientation of the listener in the virtual space. The consequence of this is that the localization of a sound source is made more difficult, especially if the reverb is too loud in relation with the direct sound and the reflections. We used relatively soft reverb in our game.

4. OVERVIEW OF THE GAME

4.1. Story

The game "Fire in Neptune" is based on the following story: The protagonist Steffen Vogel regularly visits the casino "Neptune". One night, he loses a considerable sum of money, he gets drunk and wakes up at home the next morning. During that night, the casino did burn down completely. Steffen is arrested and questioned for suspicions of arson. The recordings of the surveillance cameras in the casino have been destroyed. However, the surveillance camera of an adjacent building shows Steffen leaving the casino and driving away hurriedly before the fire breaks out. Nobody else survived the inferno.

He says time and again that he is innocent, but he cannot remember the exact course of the evening. He declares himself willing to support his statement by the "interactive memory protocol" (IMP). This is a device that measures the brain currents and provides access to lost memories. As a result, the previous evening can be reconstructed. This is where the first scene of the game begins. Steffen is instructed in the functioning of the IMP and travels virtually into the past to live through the evening.

In the game, the player has the opportunity to move freely. The first chapter takes place in the casino. Here, the player can play with the game machine or at the roulette table. In addition, he can get drinks at the bar, listen to the songs at the jukebox or listen to the conversation of other casino visitors.

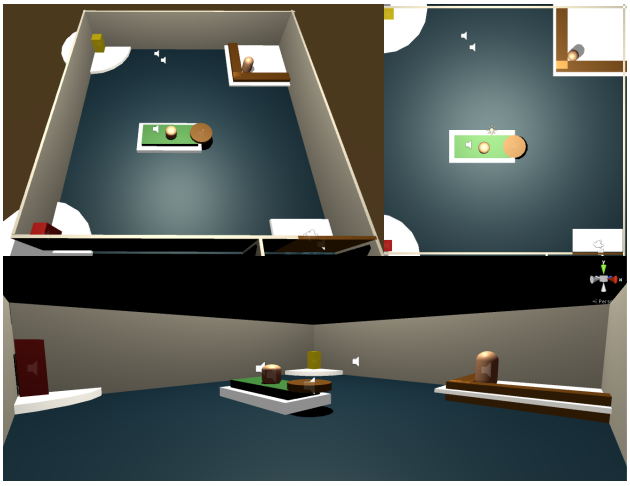


Figure 1: The layout of the casino. The 3D environment was used for development only and is not shown to the users of the audio game.

4.2. Game control

The player can move freely in the virtual space using the mouse to control the viewing direction and the keys ”W, A, S, D” to move forward, left, back, and right. This is a widely game control interface, which is used in many first-person shooter and adventure games. If a question is put to the protagonist Steffen, the player can affirm with the ”E” key or negate by pressing the ”Q” key. We limited the freedom of the viewing direction to the horizontal pane. This is different from games with similar controls, where you can also look up and down. This makes it easier to localize sounds.

Fig. 1 shows the casino. The user enters and leaves the casino through the door at the lower right. The roulette table is located in the center of the room, the bar at the top left corner, the jukebox at the top right corner, and the playing machine at the bottom right corner. The 3D environment was used for development only and is not shown to the users of the audio game. The game was developed using the Unity game engine for 3D modeling and game logic. We used dearVR to create spatial audio (see Section 3.2).

5. OBSERVATIONS

We employed an iterative design process. Six persons were regularly observed during the development of the game. The test persons differed greatly in their experience in playing computer games. All participants were without visual impairment.

The players were confronted with the game while we observed the interactions taking place and any specific issues that arose. In doing so, no help or background information was provided to the players. In order to understand the player’s actions in the virtual world, we duplicated the audio signal and listened with our own headphones. Because we had an exact idea of the structure and architecture of the rooms, we were able to ”see” exactly where the player was at any time. So we could, e.g., perceive the differences between the desired, planned behavior of the player and his or her actual actions. That way we were able to identify many problems ourselves. The players often reported difficulties later on.

5.1. Orientation at Sound Sources

Each of the test players showed the same behavior at this point: When placed in a room where only one audible signal was heard, the player oriented and walked toward it. If at a certain point in the game, there is nothing to be heard, then there is no guiding point to which the player can orient himself/herself or to determine how far he has turned in which direction. In this case, the players responded in a confused or uncertain way. Even if the sound is threatening and players move away from it, the orientation and positioning in the virtual space is always related to a sound.

In practice this means that the game developer must place a sound source at every location the player is supposed to move to. At the tutorial in the beginning of the game, the player hears a voice giving him/her the instruction to go through the door behind the voice in the wall. As soon as this sentence is pronounced, the voice is silenced and the player has to go through a door. The door is marked with an earcon that plays continuously. This sound is modeled as a soft, pulsing sound. The sound continues to play until the player has reached the door and the next sequence is initiated.

In 2002, Andersen tried out a similar approach [18], marking exits in a virtual room with distinct earcons. Unfortunately, the players were not able to target the exits. We did not face these problems in our game. We presume this could be due to the fact that the audio signal in Andersens game was limited to stereo without HRTFs, so signals from the front or from behind would sound identical.

Andersen realized that the player did not have to hear the exit but the space behind it [18]. Thus, dampened sound sources from the room behind the exit could be heard when the player passed the opening to the other room. Additionally, this solution is less artificial than using earcons and may therefore provide a better immersion. Also the player is already prepared for what he will experience in the next room.

In general, the natural sound of an object should be preferred. However, there are always situations in which the object in question does not sound in reality. Then the game developer does not have another option than to associate an artificial sound [19].

5.2. No absolute spatial orientation

For a sighted person, it can be difficult to imagine an unknown space without the help of his/her eyes. This was observed in all the players and many verbalized this fact. In most cases, the players moved from one audible object to the other without being able to make a precise statement about their absolute position in the room. For example, the player could say, ”When I go to the jukebox, I’ll first move to the bar.” But not: ”The jukebox is in the opposite corner of the room.” See Fig. 2.

Some players were completely disorientated and unable to assess whether two objects are directly adjacent or very far apart. Others could at least remember the routes that they traversed. We also observed that after some practice, users moved more efficiently and in a more goal-oriented manner (see Section 6).

5.3. Collisions with Obstacles

To orient oneself in a dark room, one does not rely exclusively on hearing but also on proprioceptive and tactile senses. It is, for example, common to advance in the dark with extended arms in order to avoid or mitigate collisions with walls or other obstacles.

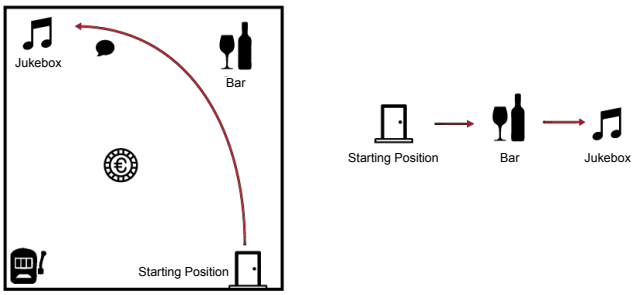


Figure 2: Left the actual route, right the perception of the player.

Lacking those modalities, collisions must be communicated acoustically in audio games.

The game "Slender - Lost Vision" (see Section 2.1.3) solves this problem with a knock sound that plays when the protagonist collides with an object. Unfortunately, the knocking sound is played as a mono sound. The knocking thus indicates that the player has encountered an obstacle but does not make any statement as to the angle at which the collision occurred. Therefore, there is no information available to the player on the basis of which he/she could opt for a course correction to the right or left.

Andersen went a step further. The signal used to indicate the contact with the wall uses stereo to distinguish whether the collision with a wall took place frontally or at an angle [18]. In our game, a knock is heard when the player hits an obstacle. In contrast to "Slender Lost Vision" and similar to Andersens game, this knocking can be heard from the direction in which the collision occurred.

We observed that some test users avoided to leave the immediate proximity of the wall and thus did not walk to the middle of the room. An almost empty space provokes this behavior much more than one in which many sound sources can be heard. Objects placed in free space are generally harder to find than those located along a wall. A possibility to break that "behavioral pattern" is to place disturbances along the room's walls.

In our game we placed these objects adjacent to a wall, which should be discovered quickly. The elements, which the player should only explore at a later stage of the game, were placed in the middle of the room.

6. EVALUATION AND DISCUSSION

6.1. Method

The players received no information or instructions on the game. However, if a player got stuck, he or she could ask for help.

The game begins with a short tutorial, in which the player is instructed how to control the game. Then, in the next section, the background story of the game is introduced. After the intro, the player is located in the casino and is confronted with an uncommented soundscape. From this point on the player takes full control. After finishing the first chapter of the game, the players rated 13 statements about their impression of the game on a 6 point Likert scale and made a sketch of the floor plan from memory.

Subsequently, the players entered the virtual casino a second time and got the instruction to move to the game machine, the bar and the jukebox in that order. The time required was recorded for each of the stations. When the player arrived at the jukebox, the final position of the route, the player was taken back to the starting position. The players then ran the same route a second time, then a third time while the time was being measured again. After that, the players were given the opportunity to correct the first sketch and make a second memory sketch. They were asked to describe in their own words what they liked most about the game and what could be improved.

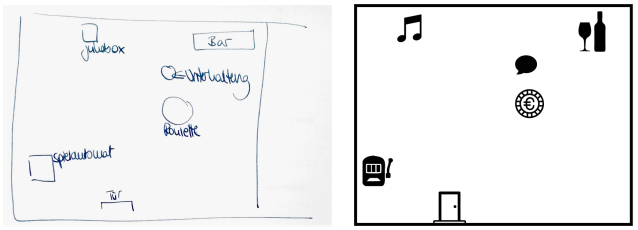


Figure 3: On the left the original plan drawn by the player, on the right the coded version.

The evaluation was conducted by ten test players and lasted about 35 minutes per player. The age of the players is between 18 and 28 years, all had unimpaired sight. Three of the players said they were inexperienced in dealing with ego perspective games. The memory sketches, which the players made, were coded, as shown in Fig. 3. If necessary, the drawing was rotated so that the door was placed at the lower edge. In this way the individual sketches could be compared with each other better than in their original form.

6.2. Results

6.2.1. Game Play

The players spent an average of 13 minutes in the game. 18% of the time was spent in the tutorial, 16% in the intro, 47% in the casino and 19% in the final navigation test, which also took place in the casino.

Firstly, it is noticeable that all 10 players state that it was a new experience for them to play an audio game (6). They liked the game (5.2) and they would gladly continue to play it (5.6). They found the story to be interesting (5.5) and the game concept to be easy to grasp (5.4). When audio games are discussed in literature, visually impaired players are often stated as the primary target group. Since this group is comparatively small, it is understandable that big game companies produce only few audio games. In 2007, blind players expressed their dissatisfaction with this situation [1]. Our questionnaire shows that players with healthy sight can find audio games interesting. Perhaps the target group for audio games is larger than generally assumed.

The players did not always know what they had to do (3.8). The individual players disagreed with each other with values ranging from 2 to 5. Probably this is a question of the playing style: While some players stopped and waited for new instructions or something to happen, others continued to explore the environment and came into action independently.

Both in positive and negative criticism, the players noticed different aspects. One player pointed out that at first it was difficult for him to distinguish between a sound source in front of and behind him. This is consistent with the use of generic HRTFs. This problem could be solved by a personlized HRTF, such as proposed

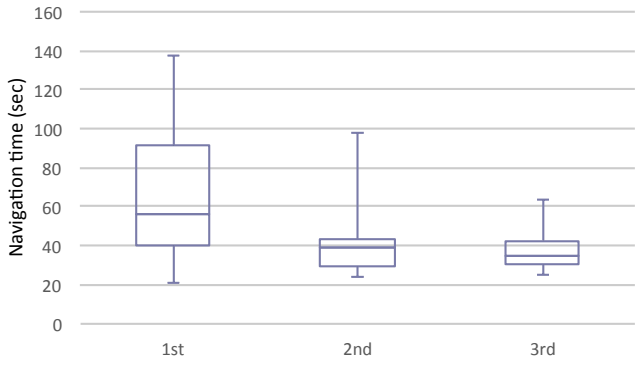


Figure 4: The time required to complete the 1st, 2nd and 3rd pass through the casino. The box plots show the minimum, the 25% quantile, the median, the 75% quantile and the maximum.

in [20].

The players took pleasure in the dialogues and the possibility to control the course of the game. In addition, the control was described as "simple, but appropriate" and the principle of navigating through hearing was described as a "new experience" and "very interesting". In addition, the the "beautiful" or "memorable sounds / voices" and the story of the game were also mentioned positively. One player liked that the scene is created in the imagination of the player, making it individual and interesting.

6.2.2. Orientation and Navigation

The players said the voices were easy to understand (6.0) and the sounds were easy to recognize (5.2). In addition, the control was described as intuitive (5.1). Although they were able to determine relatively well from which direction a sound came (4.2), they report that navigation in the room was not that easy (3.5) and that they could not imagine the space well (3.2).

There was a clear difference between the players with and those without experience in ego perspective games. While the experienced players used mouse and keyboard, the inexperienced players navigated almost exclusively with the keyboard. This resulted in a motion pattern composed of straight lines and 90 degree angles. While these players also estimated their own perception of the space as inaccurate (3.3) and felt even more unsure about their position (3.0) than the other players (3.7), their memory sketches were much closer to the actual layout of the room.

We wanted to measure how navigation time changes with playing experience. In fact, the average time required to complete the given route (from the door to the game machine, then to the bar and finally to the jukebox) reduced to 63% respectively 57% of the initial amount in the second and third run. Figure 4 provides an overview of the measured times.

We assumed that the user got more efficient in navigating with game experiences. The average time to complete the path decrease from the first time they traversed the path (67.3 s), to the second (42.4 s) and the third time (38.2 s). To examine statistical significance, we used one-way ANOVA with post-hoc Tukey Honest Significance (HSD) Test. The ANOVA results show that the time differs significantly between the three rounds ($p = 0.043$). The post-hoc Tukey HSD test did however not reach a significance level of $p < 0.05$ for any pairwise comparison. When comparing between the first and third round, significance was almost achieved with $p = 0.052$. Probably, significance would eventually be reached with more participants.

7. LESSONS LEARNED

From our experiences we want to formulate two design ideas:

- Use a sufficient number of unique sound sources.
- Use a simple floor plan.

Sufficient number of unique sound sources: As already stated, the players usually localize a sound source, orient themselves and then move towards in a more or less straight line. Once the player has reached the source, there has to be at least one additional meaningful sound source otherwise an *acoustic dead end* has been reached. Also, the sound sources should be unique. In an early version of the game, we had 12 virtual loudspeakers installed on the ceiling of the casino playing the same music. Since the sound signals were the same, the virtual loudspeaker setup did not help to navigate. Therefore, we replaced the loudspeakers with a jukebox located in one corner of the virtual room. As Andersen points out, the sounds of footsteps could be used to give the player more information about the room [18], e.g., when walking over a carpet or over stone.

Simple floor plan: The orientation can also be simplified by keeping the architecture simple [18]. It seems that many players assume that the playing area is rectangular. As a game developer, this assumption can be used in two ways. On the one hand, the space can be intentionally created in such a way that the player's expectations are met. On the other hand, it may be appropriate to deliberately make navigation more challenging. In such cases, the space may be L-shaped, octagonal or star-shaped.

8. REFERENCES

[1] R. van Tol and S. Huiberts, "What blind gamers want the industry to know. compiled for the 2006 game developers convention," 2006, last access: Feb. 2017. [Online]. Available: http://captivatingsound.com/what-blind-gamers-want-the-industry-to-know/

[2] M. J. Wolf, "Genre and the video game," in *The medium of the video game*, M. J. Wolf and R. H. Baer, Eds. Austin: University of Texas Press, 2001, pp. 113–134.

[3] T. Lokki, M. Grohn, L. Savioja, and T. Takala, "A case study of auditory navigation in virtual acoustic environments," in *International Conference on Auditory Display (ICAD)*, 2000, pp. 145–150.

[4] M. Gröhn, T. Lokki, and T. Takala, "Comparison of auditory, visual, and audiovisual navigation in a 3D space," *ACM Trans. Appl. Percept.*, vol. 2, no. 4, pp. 564–570, Oct. 2005.

[5] L. Picinali, A. Afonso, M. Denis, and B. F. Katz, "Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge," *International Journal of Human-Computer Studies*, vol. 72, no. 4, pp. 393–407, 2014.

[6] A. Afonso, A. Blum, B. F. Katz, P. Tarroux, G. Borst, and M. Denis, "Structural properties of spatial representations in

blind people: Scanning images constructed from haptic exploration or from locomotion in a 3-d audio virtual environment," *Memory & Cognition*, vol. 38, no. 5, pp. 591–604, 2010.

[7] R. H. Lorenz and H. Schuster, "Auditory pointers," in *Works in Audio and Music Technology*, A. Berndt, Ed. Dresden, Germany: TUDpress, 2015, pp. 1–30.

[8] J. Engdahl, "An evaluation of 3D sound APIs," Bachelor Thesis, University of Karlstad, Sweden, 2002.

[9] R. Lerch, G. M. Sessler, and D. Wolf, *Technische Akustik: Grundlagen und Anwendungen*. Springer-Verlag, 2009.

[10] P. Meier, "Studioakustik," in *Handbuch der Audiotechnik*, S. Weinzierl, Ed. Springer Science & Business Media, 2008, pp. 267–311.

[11] W. Ahnert and H.-P. Tennhardt, "Raumakustik," in *Handbuch der Audiotechnik*, S. Weinzierl, Ed. Springer Science & Business Media, 2008, pp. 181–266.

[12] J. Blauert and J. Braasch, "Räumliches Hören," in *Handbuch der Audiotechnik*, S. Weinzierl, Ed. Springer Science & Business Media, 2008, pp. 87–121.

[13] W. Mueller and F. Ullmann, "A scalable system for 3D audio ray tracing," in *IEEE International Conference on Multimedia Computing and Systems*, vol. 2. IEEE, 1999, pp. 819–823.

[14] N. Röber, U. Kaminski, and M. Masuch, "Ray acoustics using computer graphics technology," in *10th International Conference on Digital Audio Effects (DAFx-07)*, 2007, pp. 117–124.

[15] N. Röber, M. Spindler, and M. Masuch, "Waveguide-based room acoustics through graphics hardware," in *International Computer Music Conference (ICMC)*, 2006, pp. 21–28.

[16] J. Dvorak, C. Pirillo, and W. Taylor, *Online!: The Book*. Prentice Hall Professional, 2004.

[17] dearVR, "dearVR 3D audio reality engine – user manual v1.1," 2016, last access: Feb. 2017. [Online]. Available: http://www.dear-reality.com/support/dearVR_user_manual.pdf

[18] G. Andersen, "Playing by ear: Using audio to create blind-accessible games," 2002, last access: Feb. 2017. [Online]. Available: http://www.gamasutra.com/resource_guide/20020520/andersen_pfv.htm

[19] M. A. Oren, C. Harding, and T. Bonebright, "Evaluation of spatial abilities within a 2d auditory platform game," in *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2008, pp. 235–236.

[20] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha, "P-HRTF: efficient personalized HRTF computation for high-fidelity spatial sound," in *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 53–61.

EVALUATING TWO WAYS TO TRAIN SENSITIVITY TO ECHOES TO IMPROVE ECHOLOCATION

Laurie M. Heller, Arley Schenker, Pulkit Grover, Madeline Gardner, Felix Liu

Carnegie Mellon University
Pittsburgh, Pennsylvania

ABSTRACT

We investigated whether training sighted individuals to attend to information in echoes could improve their active echolocation ability. We evaluated two training techniques that involved artificially generated sounds. Both artificial techniques were evaluated by their effect on natural echolocation of real objects with self-generated clicks. One group trained by discriminating sounds presented over headphones in the lab. The lateral displacement or distance of the echo was varied in a staircase procedure. The second training group used an echolocation app on a smartphone. They navigated a maze by using echo cues presented over earbuds. The echo cues had 3D audio virtual reality cues. Participants in the control condition did not improve but the majority of participants who trained did improve. The lab training is labor intensive whereas the app training was self-guided and convenient. This has implications for training methods aimed at echolocation that might ultimately be useful for navigation by visually impaired individuals.

1. INTRODUCTION

Visually impaired people use sound and other sensory cues to navigate, compensating for their loss of vision. In rare cases, they develop the skill of echolocation, in which users self-generate tongue clicks in order to gain information about the surrounding environment. These source, or referent, clicks reflect off of surrounding objects and travel back to the ears, allowing echolocators to perceive information about the location and characteristics of the reflecting object. Early research on echolocation most often concerned bats and dolphins; however, there is a recent increase in effort to study human echolocation and other navigational techniques to inform navigation devices for the visually impaired population [1].

Echolocation using tongue clicks does not occupy a hand (as does a cane, guide dog, or most other navigational devices). The only devices necessary are the head, mouth, and ears, so echolocation is neither expensive nor unwieldy. Echolocators have substantial control over their own tongue clicks. They can also move their head in order to emit clicks in varying directions, which allows them to more precisely detect objects on their left or right. For these reasons, training to use echolocation has the potential to help visually impaired people with normal hearing to navigate more independently. Current systems, some using auditory virtual reality, exist in laboratories for echolocation study purposes [2], [3]. However, it is desirable to make echolocation training easier and more accessible. For this reason, we assessed two methods for training people to discriminate artificial echo information and we measured its success with natural echolocation. We used sighted participants for this preliminary test in order to refine the methods.

2. ECHO ACOUSTICS AND PSYCHOACOUSTICS

Echolocation, while a potentially useful skill, is not in common use as a navigational tool. First of all, echoes are subject to masking in loud environments as echoes are very low amplitude sounds. They are acoustically complex and depend on many factors in the physical environment such as how reflective the surrounding material is. Second, echoes are difficult to use because humans normally suppress the locations of environmental echoes.

2.1 The precedence effect

First described by Wallach and colleagues, the *precedence effect* describes the localization of two sounds occurring in short succession. When two similar sounds with a small delay between them are played from different locations, both are heard to have come from the location of the first sound [4]. This gives precedence to

https://doi.org/10.21785/icad2017.053

the location of the first sound. As an echo is simply a quieter version of a referent click, the precedence effect is detrimental to echolocation. People can be trained to be more sensitive to echo information after many hours of training in a laboratory task [5]. The precedence effect has been quantified in an echolocation context by Wallmeier and colleagues, where participants localized echoes reflected by a single sound reflector and the leading and lagging of two reflectors [6].

2.2 Acoustic cues to location

When a click is emitted, it bounces off the objects and surfaces in the environment that do not absorb it. Many echoes return to the ears, and several computations can occur to determine information about the location of the sound. For instance, the closest object to the observer will return the first echo. The time between the outgoing (referent) click and the echo will be directly related to the distance of the reflecting object and the speed of sound. Although echo delay is a reliable cue to distance, there are other less reliable cues to distance such as frequency content or echo level relative to the source level. Sound loses intensity the farther it travels [7]. These cues can be simulated by modeling the physical environment, including the absorption characteristics of all surfaces. For the lateral and vertical localization of echoes, there are even more cues. A sound coming from the left side of the listener, for example, arrives at the left ear before arriving at the right ear [7]. These interaural time differences (ITDs) are on the order of microseconds. As sound travels it can also lose intensity at one ear relative to the other, depending on the frequency and direction of the sound. These interaural level differences (ILDs) are another source of information used to localize sound information. Furthermore, individuals' outer ears, or pinnae, are uniquely shaped to amplify certain frequencies and attenuate others, depending on the direction of the sound source. Both the head and pinna come together to form a complex direction-dependent filter [7]. This filtering of incoming sound is known as a head-related transfer function (HRTF). These computations must occur simultaneously for all incoming sounds, and in a reverberant environment, it can be difficult to localize the various overlapping sounds.

2.3 Training to use echo information

Surprisingly, given all these acoustic challenges, both blind and sighted individuals can be trained to use echoes to glean information about objects around them. Studies comparing the two populations often compare blind expert echolocators, blind non-echolocators, and sighted individuals. Teng and Whitney showed that after four hours of training, sighted participants could use self-generated clicks to discriminate the size of an object as well as a blind expert echolocator could [8]. Additionally, Schenkman and Nilsson showed that blind participants were better able to report a sound that had been recorded in the presence of a reflecting object, compared to sighted participants; however, both blind and sighted participants could perform the task after training [9]. Finally, sighted participants have been shown to use echoes to discriminate the distance of objects after only one hour of training [10].

Although it has been shown that both blind and sighted individuals can perform simple echolocation tasks, the majority of them have been trained under the same conditions as the test. Even though lab training for specific focused tasks can succeed in a few hours, the training process for individuals who use echolocation on a daily basis for navigation can take years. We therefore searched for a way to make the initial learning process easier and less task-specific. We hypothesized that individuals would be able to learn to discriminate echoes if they were enhanced. As echolocation has been shown to have functional benefits for its users, the process by which individuals become proficient echolocators would ideally be as easy as possible [11]. We amplified the echoes above realistic levels and we decluttered the sound environment. In this study, we investigated whether asking sighted individuals to discriminate these artificially enhanced echoes in a variety of discrimination tasks could help them to actively echolocate large objects using natural mouth clicks. We used two methods for training, one involving a smartphone application that was a game called EchoExplorer™ (described in more detail in [12]). The other training method was the traditional psychophysical technique of training using a staircase adjustment method while doing two-interval forced-choice discrimination of both distance and lateral position.

3. METHODS

3.1 Design

Thirteen naive sighted participants were tested, six in the App condition (average age 21.8), five in the Lab condition (average age 19.1), and two in the Control condition (average age 20.5). All participants were first

pre-tested on real-world discrimination of board positions using their own clicks for echolocation. Next, the Lab and App participants completed 15 hours of echo discrimination training prior to being post-tested in the same task. The Control group did no training but waited a similar amount of time between pre and post tests. The methods of training for the App and Lab groups were different and are described in separate sections below. Written consent was obtained from all participants, participants were paid, and the study protocol was approved by the Carnegie Mellon University Institutional Review Board (IRB).

3.2 Echolocation pre and post-tests

In the pretests and posttests we asked participants to indicate the locations of objects but did not give them any feedback; additionally, we discouraged outside research on human echolocation between pre and post tests. During pre and post-tests, blindfolded participants made clicks to locate a board while seated in the center of a large, otherwise empty room. We told them how to make palatal clicks but gave no training or feedback. Across trials, a large board was located in one of 4 lateral angles and one of 3 distances. Participants localized boards with either a fixed head position or with free head movement.

The ½" thick melamine board measured 24" by 48" and was reinforced throughout its length by a 2x4" wood beam to enhance its rigidity. The board was held vertically and a 1" foam pad was added to its bottom edge to help reduce the sound produced by setting the board down. The board was placed in one of 12 locations relative to the front of the participant's head: one of three distances (0.9, 1.8 or 2.7 m) and one of four lateral angles (45 or 90 degrees to the right, 45 or 90 degrees to the left). Two trials at each possible location yielded 24 test trials that were presented in one of four counterbalanced random orders. In order to mask any subtle sounds made by experimenters placing the board, participants listened to white noise through earbuds during trial set-up and take-down. At the start of each session, the white noise level was set empirically to be at a comfortable level that successfully masked the sound of experimenter asking the participant a question at a normal conversational level.

The participant was blindfolded while seated in the center of an unfurnished 28' x 56' carpeted room with fabric-covered walls. The experimental trials were preceded by one practice trial. At the start of a trial, the participant placed earbuds in their ears and listened to white noise. The experimenter stood at a central starting location, 2.7 meters directly in front of the participant (holding the board) and announced loudly that the next trial was starting. At that point the listener began counting silently up to 10. During that time, the experimenter with the board moved the board into position and stood behind the board. The participant stated "ready" before taking their earbuds out. Holding their head level and facing directly ahead, they made 10 clicks with their mouth. They then guessed the direction and distance of the board (e.g. 45 right, 2 for medium distance). Immediately after this, they freely moved their head to the left and right while clicking 10 more times. After this, they made another estimate as to the board's direction and distance. Participants put the earbuds back in, listening to noise while the board was moved back to the center starting location before the next trial began.

3.3 Game training

3.3.1 Participants

After pretesting on echolocation, six naive users were trained using a beta version of the EchoExplorer™ app [12] for 15 hours. This game was an app on a smartphone that was designed for this purpose and could be played at the participant's convenience. This training duration was chosen because video game experiments suggest that 15 hours is sufficient to induce some measurable changes in transferred skills [13]. All of these participants had normal hearing according to in-lab audiograms performed with a MAICO MA41 audiometer.

3.3.2 Procedure

The goal of the game was to provide a training platform for learning echolocation. We designed a game that requires the user to navigate through various mazes using simulated echoes. An avatar is used to represent the current location of the user in the maze. Although it is possible to see a depiction of the maze and the avatar in a debugging mode, *all game play was done with a blank screen* (no visual information). At any point in time the user can instruct the application (e.g. by tapping on the touchscreen) to generate echoes based on the current location of the avatar within the maze. The application generates a click followed by the echoes that convey spatial information through echo delays, ITD, and/or ILDs (see Section 2). For instance, if the user is facing a close wall straight ahead, the referent click and the resulting echo will be heard in quick succession. If on the other hand, the wall is much further from the

user's current location, the difference between the time when the reference click is heard and when the corresponding echo is heard will be much larger.

We used the high-resolution acoustics modeling program Odeon, a room acoustic simulation and measurement software, in order to artificially create echoes of clicks based on the room shape. This program generates acoustic cues based on the shape of the maze: for example, the time delay between the emitted sound and the echo provides a reliable cue as to the distance of nearby obstacles. The spatial location of the nearby obstacles is further indicated by the acoustic effects of the head in the path of the echoes. (HRTF 58 was used from the CIPIC database [14] because it was measured from a head with anthropometric feature closest to the average of the CIPIC database and has been validated with naive users in favorable comparison to KEMAR [15]. However, combinations of HRTFs may have different acoustic properties than individual HRTFs and may be worth using in the future [16]).

After creating a set of echoes in Odeon to simulate hallways ending in all the possible junctions (deadend, elbow left, elbow right, t-junction, and stairwells) we boosted the level of all the echoes by the same amount to increase their audibility while keeping the outgoing reference click at its original level. This helped users distinguish between the subtle changes in echoes as they moved through a hallway. After pilot testing, the echoes were boosted 21 dB at the start of the game. Every 15 levels in the game, the echo boost was decreased by 2 dB until users indicated that the game was too difficult to play. This lower limit occurred at or above an echo boost of 7 dB or all users.

Auditory cues were given after each move gesture just to indicate that the motions had been accomplished, but no correct/incorrect feedback was given for individual motions unless that motion caused the avatar to crash into a virtual wall of the maze, in which case, a "crash" sound was played. Implicitly, the absence of a crash sound after a motion indicated that it was a safe move, but nothing indicated whether or not the avatar was heading in the correct direction. However, if the avatar returned all the way to the start of the maze, an auditory warning was played.

3.4 Traditional Psychoacoustic Training

*3.4.1 Participant*s

After pretesting with echolocation, we trained 5 sighted college students in the laboratory for 15 hours with artificial sounds. All participants were verified to have normal hearing except for one participant who had a mild low-frequency hearing loss in one ear (25 dB HL) from 250 to 1000 Hz; who exhibited normal sensation levels and normal interaural discrimination thresholds using our broadband click stimuli.

3.4.2 Stimuli

Palatal mouth *referent clicks*, recorded by an undergraduate research assistant, were used as a foundation for the stimuli. These clicks were recorded using Roland CS-10EM binaural microphones in an empty IAC sound-attenuating chamber in which the walls and ceiling were covered with 4-in. Auralex foam wedges to reduce echoes. Over 50 clicks were recorded, but only the 17 clicks with waveforms similar to the ideal palatal clicks described by Rojas were used [17]. A custom echo generation program, written in Matlab, was applied to each of these clicks to generate realistic echoes. The echo was generated by adding a copy of the referent click at the appropriate delay corresponding to the distance of the reflecting object. The referent sound travels to the reflecting object and back to the listener as an echo, so it travels twice the distance between the listener and the object. Then, using a speed of sound of 343 m/s, an echo from an object 5 m away, for example, would occur 29 milliseconds ((2 * 5) m / 343 m/s) after the onset of the referent.

When tracking on lateral position, ILDs were implemented in the left and right channels by attenuating one channel according to the angle and distance of the reflecting object. The maximum ILD, used when an object was 90º to the left or right, was 10 dB. ILDs for angles between 0º and 90º were ((1/9) * angle) dB. The overall level of the echo decreased in both channels by an additional 6 dB each time distance was doubled relative to 1m. During some tracks interaural time delay (ITD) was manipulated instead of ILD. (Those data obtained with ITDs are not presented here due to incomplete data sets.) Determination of the echo levels at which the tasks could be performed at a medium difficulty level for the average person occurred during pretesting. In subsequent staircases, echo levels were reduced as needed to keep thresholds similar over time. Echoes were generated at 10-degree intervals between -90º (90º to the left) and +90 (90º to the right). In addition, echoes were generated at distances between 1 and 5 meters at 0.5-meter increments. The referent click was always presented at the same level, 30 dB SL (sensation level). Within one trial, the reference click was the same for both intervals; the only difference

between the intervals was the timing and level of the echo, calculated relative to the referent click, that accompanied each click. Between trials, clicks were chosen randomly from the 17 clicks described previously.

3.4.3 Procedure

After providing written consent, participants were informed about the structure of the tasks and were given the opportunity to ask questions. They were also instructed not to focus on any one cue in the stimuli and to close their eyes during the experiment. They performed the experiment while seated at a computer in the aforementioned listening chamber. Participants listened to the stimuli through Sennheiser HD 600 headphones.

There were two main tasks: distance and left/right discrimination. All participants performed the distance task first within each of the 15 hour-long test sessions. Each of these tasks was a 2-interval forced choice (2IFC) task in which the order of the intervals was randomized. During the distance task, each interval of a trial contained a referent click, which was centered (no ILDs were applied) and whose channels were normalized relative to each other, followed by an echo generated according to the aforementioned parameters. The angle of echoes in the distance task was kept constant at 0º. Participants discriminated a click with a close echo from a click with a far echo and reported which interval, 1 or 2, contained the closer echo by pressing the corresponding key. During the left/right task, the echo always had a simulated distance of 1 m whereas the angle varied between -90º and +90º at 10-degree increments. In the left/right condition, participants reported whether the echoes in the two intervals moved from right to left or from left to right by pressing the 1 or 2 key, respectively.

During each trial of the 2IFC tasks, participants were shown a sentence reminding them of the correct key presses. (For example, in the distance condition: "Which click contained the closer echo? (1 2)".) A trial consisted of two clicks separated by 500 ms of silence. For example, a trial in the distance condition could contain a click with an echo from 1 m away, followed by 500 ms of silence, followed by a click with an echo from 4.5 m away. Prerecorded verbal feedback was given (e.g. "correct!"). The next trial was determined using a three-down-one-up staircase paradigm [18]. The staircase paradigm allowed for the determination of a threshold at which participants responded accurately to about 78% of the trials. At the beginning of all three conditions, the staircases started at the easiest level (distance: 5m, lateralization and left/right: 90 degrees). If participants correctly answered three trials in a row, the subsequent trial increased in difficulty by one level. If participants incorrectly answered a single trial, the subsequent trial decreased in difficulty by one level. The track ended after 11 reversals were observed or if the track lasted for over 70 trials without 11 reversals. Here, reversals are defined to be points during the track where participants answered correctly three times after answering incorrectly on the previous trial, or points where participants answered incorrectly once after answering correctly on the previous trial. Participants performed 2-5 tracks per condition per test session. A condition average for each participant was calculated by computing the average of each of that participant's tracks. If the track did not contain 6 reversals or if the participant performed more than 3 trials at the easiest level during any given reversal in the last 6 reversals, that track was not included in the participant's threshold calculation. The staircase adjusted the distance (in depth, or angular distance) between the two intervals of a trial and determined a threshold. After each track, performance was reviewed by an experimenter. If the participant's performance was good, the level of the echo was decreased in the next track. In this way, the experimenter aimed to keep the threshold relatively constant while gradually decreasing the echo level over time.

4. RESULTS OF GAME PLAY VS. PSYCHOACOUSTIC TRAINING

4.1 Learning during training

Using the EchoExplorer™ game, we measured number of crashes into walls per level, number of echoes requested per level, number of steps taken per level, and active time per level. Because echo level did decrease by 2 dB every 15 levels after the tutorial, in order to look for learning effects, we pulled out performance at a few echo levels to compare the number of crashes per level as training went on. The maze level was cycled through a few times so that we were able to compare performance at similar echo/maze levels over time. Figure 1 shows log fits to the average number of crashes per level as a function of the sequential time of play of each level. Crashes were higher initially for the earliest level with the 21 dB boost because it was the first level encountered but, as expected, the asymptotic

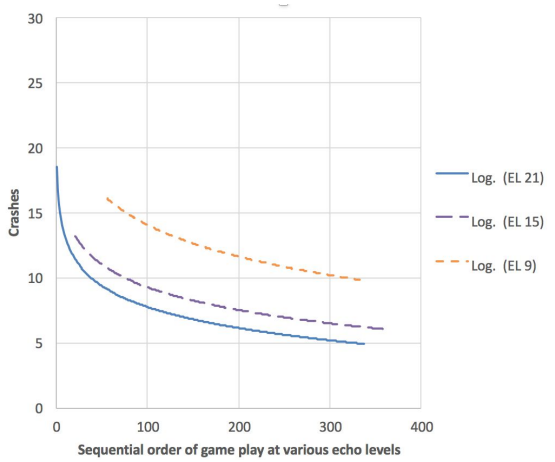


Figure 1. Number of times crashing into maze walls as a function of the sequential order in which the level was completed. The parameter is the level boost of the echo in dB, either 9, 15 or 21 dB. The average data across all App participants are fitted with Log functions.

performance level was best for this condition. Even at the most difficult levels, the echo was still boosted beyond what would typically occur in a real hallway.

There was evidence of learning during the lab training. Across participants in the Lab group, the average simulated distance that supported 78% discrimination from a 1m distance was 4.45 m (SD 0.9). Figure 2 shows the echo level required for discrimination of distance as a function of training hour for four participants. (Varying echo level data were not available for the fifth participant due to a procedural error). Improvement (measured as a decrease in echo level) ranged from 4 to 12 dB over time. Across participants, the average simulated angle difference that supported 78% correct discrimination between right and left lateral positions was 38.4 degrees (SD 9.9). Figure 3 shows the echo level required for discrimination of lateral position (in the left/right task) for all 5 participants in the Lab group. In both graphs, the echo level required decreased over time. The echo was still boosted beyond what would typically occur in natural conditions, but by the end, all participants could reliably discriminate the echo when its level was lower than the source level.

4.2 Improvement in echolocation

Discrimination of lateral position and distance was evaluated in both a pretest and posttest. Dprime was the measure of sensitivity to the different possible locations of the board [19]. The four lateral positions (-90, -45, 45,

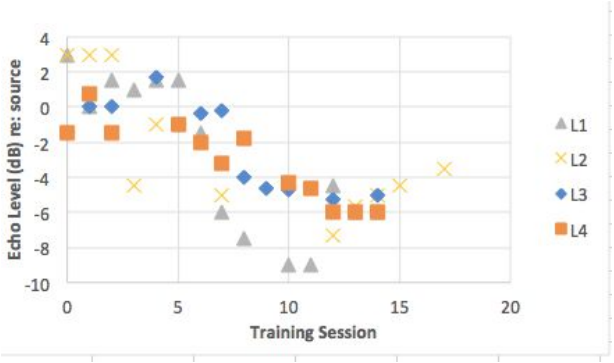


Figure 2. Echo level used relative to the initial outgoing click in order to support average threshold performance on distance discrimination as a function of hours of training in the lab (for individuals in the Lab group).

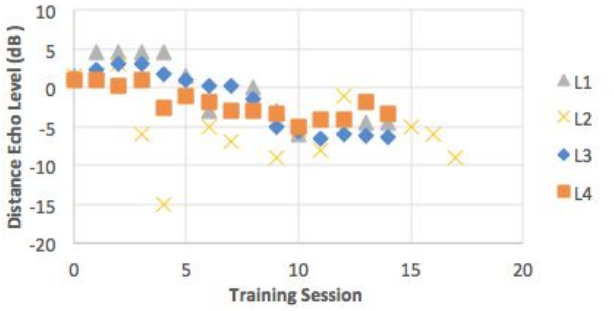


Figure 3. Echo level used relative to the initial outgoing click in order to support average threshold discrimination of left/right lateral position as a function of hours of training in the lab (for the Lab group).

and 90) yielded a chance level of 25% which would be equivalent to a d' of 0 for a four-alternative forced-choice task. The three possible distances (0.9, 1.8 and 2.7 meters) yielded a chance level of 33% with a d' of 0 for a three-alternative forced-choice task.

When head position was fixed, there was modest sensitivity to lateral position with an average d' of 0.10 at pretest and 0.43 at posttest, with 9 of 11 trained participants showing improvement. When the head was moved freely, the average lateral position discrimination was 0.15 at pretest and 0.64 at posttest, with 10 of 11 participants showing improvement. For the head moving condition, post test d' versus pretest d' is shown for all observers in Figure 4. No change in performance would be implied by the dashed line, whereas improvement is indicated by all data points above that line. The average sensitivity was relatively low given that a d' value of 1 is typically considered threshold discrimination, similar to the 78% correct

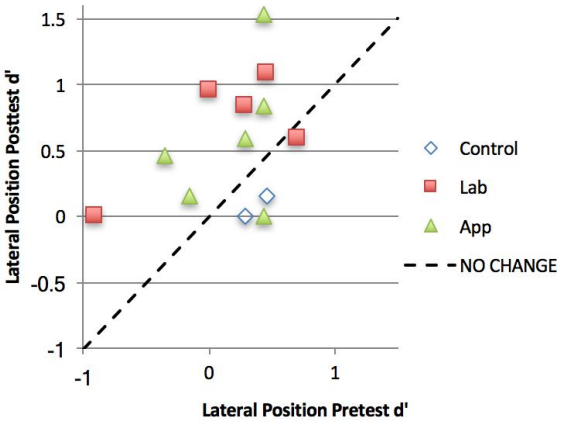


Figure 4. Post test d' (sensitivity) to lateral position as a function of Pre test d' in a real-world discrimination task when the head was moved freely. Control participants (open diamonds) did not have any training between tests. Lab participants (red squares) and App participants (green triangles) were trained in the lab, or used the app, respectively, for 15 hours between tests. Data points above the dashed line indicate improvement.

value that we targeted in the lab training. Nonetheless, the average improvement in d' (0.30 across conditions, 0.11 for fixed, 0.49 for moving) was reliably greater than zero (tested by a mixed ANOVA using a between subject factor of training group and within-subject factors of fixed vs moving head and pre vs post test ($F(1,4)=63$, $p<.001$). There no significant main effect of the type of training (Lab and App groups) and no effect of the fixed and moving head conditions, nor were there any interactions. Note that the condition in which the head moved also benefited from more observations because it came after the fixed condition, so we would not draw conclusions about head movement *per se* from this result. Among the 44% of post-test lateral position trials in which trained participants changed their answers between the fixed and head moving conditions, they changed from an incorrect to a correct answer 42.6% of the time (whereas chance would be 33%). The two control participants displayed pretest dprimes within the middle of the range of all other participants and did not improve their sensitivity between pre and post tests.

Average discrimination of distance was low in the 3afc task, but reliably above chance (average of d'=0.215, with 95% confidence intervals for all conditions above zero). There was no learning of distance between pre and post tests as indicated by an ANOVA with factors of training group, pre/post test and fixed/moving head. Average distance d' did not improve between pretest (0.22) and posttest (.21) and this did not interact with training group or head movement. Data were missing from one participant in the lab group in the fixed head distance discrimination; the resulting App group did have a lower average d' (0.12) than the Lab group (0.31) ($F(1,8)=6.13$, $p<.038$), but because distance discrimination did not significantly change at post-test, this group difference was inconsequential for characterizing training effects. Dprime for distance was not reliably lower when the head was fixed (0.19) compared to moving (0.24) nor did it interact with any other factors.

5. CONCLUSION AND DISCUSSION

At the outset, there were many factors working against the possibility that our first attempt to train people with this game would produce measurable echolocation benefits in the real world. Among these factors were the fact that this was a beta test of an app; our participants were sighted; the training conditions did not match the test conditions in terms of the reflector locations; the referent clicks were not the participant's own clicks; the HRTFs were not emitted by the participant; and the training echoes were louder than they are in real life. For this reason, we find the small reliable benefits that we measured to be encouraging.

If sighted users can gain some small benefit from this training, the next step after further refinement would be to test blind users. At this point, we can conclude that training with artificial sounds - sounds that do not require the participant to emit clicks or move - can be effective in improving echolocation with self-generated clicks in the real world. At the start of this study it was not known whether a game such as ours could produce a benefit.. Note that this new finding of transfer is distinct from showing that training on echolocation in the real world improves echolocation in the real world. Even so, locating a board while moving one's head is only one small step towards using echolocation to navigate while moving in a realistic environment. Although it is likely that our artificial training may be less effective than if the same amount of time were spent navigating in the real world while exclusively using mouth clicks, both of our training methods have potential advantages, especially when many hours of solo training are required. Our methods allow the use of enhanced echoes, and our data show that this enhancement was required to achieve reliable discrimination at the start for all

participants. We saw that learning in the lab using a standard psychoacoustic method was not substantially superior to learning with the app. However, a key advantage to training with the game is that it is under the user's control and can be used at their convenience; it therefore is more accessible and practical than the customized psychoacoustic training method in the lab. It should also be pointed out that the EchoExplorer™ game was tested as a beta version and is not yet optimized. Therefore, we find these experimental results encouraging for the future of using games to learn new ways to use sound for navigation.

6. ACKNOWLEDGMENT

We thank Sarah Kwan, Tejal Kudav, Kiran Matharu, Jacqueline Hon, Jessica Kwon, Aaron Schwartz, Chieko Asakawa, Catherine Getchell, participants at the Blind and Vision Rehabilitation Services of Pittsburgh, and Art Rizzino for helpful discussions, feedback and pointers. We also thank Google Inc., the NSF Center for Science of Information (CSoI), CMU's SURG program, and the NSF REU program for their generous support.

7. REFERENCES

[1] A. J. Kolarik, S. Cirstea, S. Pardhan, and B. C. J. Moore, "A summary of research investigating echolocation abilities of blind and sighted humans," *Hearing Research*, vol. 310, pp. 60-68, 2014.

[2] B. F. G. Katz and L. Picinali, "Spatial audio applied to research with the Blind.," in *Advances in Sound Localization*, 2011, pp. 225–250.

[3] D. Pelegrin-Garcia, M. Rychtáriková, C. Glorieux, and B. F. G. Katz, "Interactive auralization of self-generated oral sounds in virtual acoustic environments for research in human echolocation," in *Proceedings of Forum Acusticum 2014*, 2014.

[4] H. Wallach, E.B. Newman, and M. R. Rosenweig, "The precedence effect in sound localization," *The American Journal of Psychology*, vol. 62(3), 315-336, 1949.

[5] K. Saberi, and J. V. Antonio, "Precedence-effect thresholds for a population of untrained listeners as a function of stimulus intensity and interclick interval," *J. Acoustical Soc. Am*., vol. 114, pp. 420-429, 2003.

[6] L. Wallmeier, N. Geßele, and L. Wiegrebe, "Echolocation versus echo suppression in humans," *Proceedings of the Royal Society B.,* vol. 280*,* 20131428, 2013.

[7] B. C. J. Moore, *An Introduction to the Psychology of Hearing, Sixth Edition.* Bingley, UK: Emerald Group Publishing Limited, 2012.

[8] S. Teng, and D. Whitney, "The acuity of echolocation: Spatial resolution in the sighted compared to expert performance," *Journal of Visual Impairment and Blindness*, vol. 105(1)*,* pp. 20-32, 2011.

[9] B. N. Schenkman, and M. E. Nilsson, "Human echolocation: Blind and sighted persons' ability to detect sounds recorded in the presence of a reflecting object," *Perception*, vol. 39*,* pp. 483-501, 2010.

[10] A. Tonelli, L. Brayda, and M. Gori, "Depth echolocation learnt by novice sighted people," *PLoS One*, vol. 11, no. 6, pp. 1–14, 2016.

[11] L. Thaler, "Echolocation may have real-life advantages for blind people: an analysis of survey data," *Frontiers in Psychology,* 2013, vol. 4(98), eCollection.

[12] W. Wu, R. Morina, A. Schenker, A. Gotsis, H. Chivukula, M. Gardner, F. Liu, S. Barton, S. Woyach, B. Sinopoli, P. Grover, and L. M. Heller, "EchoExplorer™: A game app for understanding echolocation and learning to navigate using echo cues," in *ICAD Proceedings 2017,* 2017.

[13] C. S. Green, D. Bavelier, "Action video game training for cognitive enhancement," *Current Opinion in Behavioral Sciences*, vol. 4, pp. 103-108, 2015.

[14] R.V. Algazi, R.O. Duda, D.M. Thompson, C. Avendano, The CIPIC HRTF Database, *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics* In WASSAP '01 (2001).

[15] E. De Sena, N. Kaplanis, P.A. Naylor, T. van Waterschoot, "Large-scale auralised sound localisation expe riment," *AES 60th International Conference*, 2016.

[16] L. S. R. Simon, N. Zacharov, and B. F. G. Katz, "Perceptual attributes for the comparison of head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 140, no. 5, pp. 3623–3632, 2016.

[17] J. A. M. Rojas, J. A. Hermosilla, R. S. Montero, and P. L. L. Espi, "Physical Analysis of Several Organic Signals for Human Echolocation: Oral Vacuum Pulses," *Acta Acustica United with Acustica,* vol. 95, pp. 325-330, 2009.

[18] M. R. Leek, "Adaptive procedures in psychophysical research," *Perception & Psychophysics,* vol. 63(8), pp. 1279-1292, 2001.

[19] N.A. Macmillan and C.D. Creelman, Detection Theory: A Users Guide, 2nd Ed. Psychology Press, 2004.

Paper Session 6
Philosophy/Aesthetics

SPECTRAL PARAMETER ENCODING: TOWARDS A FRAMEWORK FOR FUNCTIONAL-AESTHETIC SONIFICATION

Takahiko Tsuchiya and Jason Freeman

Georgia Institute of Technology,
Center for Music Technology,
840 McMillan St. Atlanta GA 30332, USA
{takahiko, jason.freeman}@gatech.edu

ABSTRACT

Auditory-display research has had a largely unsolved challenge of balancing functional and aesthetic considerations. While functional designs tend to reduce musical expressivity for the fidelity of data, aesthetic or musical sound organization arguably has a potential for representing multi-dimensional or hierarchical data structure with enhanced perceptibility. Existing musical designs, however, generally employ nonlinear or interpretive mappings that hinder the assessment of functionality. The authors propose a framework for designing expressive and complex sonification using small timescale musical hierarchies, such as the harmony and timbral structures, while maintaining data integrity by ensuring a close-to-the-original recovery of the encoded data utilizing descriptive analysis by a machine listener.

1. INTRODUCTION

The long-standing dilemma in data sonification research between functional and aesthetic approaches suggests the difficulty of simultaneously achieving an accurate conveyance of information and a complexity or expressivity of the sound output. As functionalists tend to eliminate unnecessary elements in sonification while aesthetic proponents often interpret data subjectively or employ external information for a "metaphorical" mapping [1], finding a middle ground between them seems challenging. Besides such arbitrariness in mapping decisions, many musical sound organization principles, such as scaling and quantization of the pitch and time, typically require non-linear transformation of data that may result in the loss of information. Despite these challenges, music, as organized sound [2], arguably possesses a potential for multi-dimensional mapping of data optimized for human perception with unique hierarchical sound organizations. We investigate the possibility of this multi-dimensional and higher-order expression for data sonification with enhanced perceptibility while acknowledging the risk of information loss.

1.1. Musical Structures

The concepts of musical structures are extensive and differ amongst music theories, compositional styles, and musicology or music information retrieval (MIR) research. In the most ordinary case with western symbolic notation, a musical event, for example an onset, may have pitch, volume, timing, duration, and timbre as parameters. The pitch may have higher-level structures including harmony, scale, or melodic patterns, while the volume and timing may contribute to the forming of timbre or rhythm. Such hierarchical relationships are not limited to particular styles of music such as Western classical music, but, as analyzed by Roads [2], may apply with varying degree to any aesthetic sound expressions as they all derive from a single time-domain acoustic phenomena. Musical structures may also derive from, besides the level of time resolution, various non-linear data transformations such as scaling (e.g., in range and distribution), quantization, alignment of continuity or discontinuity, dichotomous balances such as repetition and variation, tension and release, and noisiness and tonality. In this paper introducing our initial version of the framework, we limit the focus of musical structures to smaller time-scale hierarchies centered around a common musical event (i.e., note) in relation to the frame-level spectral analysis employed in the assessment of data integrity.

1.2. Evaluation Frameworks

A major problem in incorporating expressive or aesthetic techniques in sonification design is the general lack of ways to ensure the data transmission in a quantifiable manner. Many existing evaluation attempts with aesthetic sonification are high-level subjective listening tests that are inevitably influenced by subjective listener preferences and by listener fatigue. In the proposal of Sonification Evaluation eXchange (SonEX) [3], Degra et al. discuss the importance of quantified measurements such as accuracy, error rates, reaction speed, and precision measures. However, even though SonEX provides a framework for standardized and reproducible experiments, it does not present cases for how to actually measure the accuracy and error rates in a complex sonification. Another evaluation scheme, called multi-criteria decision aid, proposes quantitative measurement of the design features, such as clarity, complexity, and amenity, in various sonifications [4]. These evaluation schemes put weight on defining a community-based environment for comparative testing of new sonification designs. However, there seems to be few attempts in objective measurements within a single aesthetic sonification system. While subjective listening tests and statistical analysis may reveal individual mapping patterns with perceptual increase or decrease, it is not practical to test the extensive hierarchical relationships in an expressive sonification against varying data sources. Our framework, therefore, focuses on assessing the minimization of measurable loss of information in the process of musical mapping and transformation. Instead of extending the existing qualitative evaluation methods, we propose the introduction of

https://doi.org/10.21785/icad2017.051

a machine-listening element, found in speech recognition and MIR [5][6]. Although this study does not directly address the measurement of the perceptive quality of musically-structured sonification, which is still of great importance, using the techniques employed in speech recognition or MIR allow for quantitative measurements that are often modeled around human auditory perception [7] such as tonality and noisiness measurements [6]. MIR also works directly on complex and multi-dimensional sound organizations (i.e., music), which is suitable for the musically-complex (e.g., polyphonic, spectrally rich) sonifications that we hope to achieve. Our experiments reported in this paper employ the most deterministic techniques in the encoding and decoding processes to achieve the retrieval of the data.

1.3. Sonification Frameworks

For designing functional-aesthetic sonification with an information-fidelity evaluation, we propose a framework with structural analysis and mapping as well as frequency-domain parameter encoding and decoding processes. We call this framework *spectral parameter encoding* (SPE) for musically expressive sonification.

SPE draws insights from prominent sonification frameworks such as parameter mapping sonification (PMS or PMSon) [8], model-based sonification (MBS) [9][10], and audification [11]. With many overlapping strategies with SPE, PMS provides comprehensive techniques for semi-automating (i.e., aiding decision making) the process of mapping unseen data to various audio-synthesis parameters. It presents many considerations for optimizing the data features, such as the dynamic range, for better perception and clearer auditory presentation with relatively simple mapping to common synthesis parameters. Grond and Berger also discuss the artistic applications (i.e., "musification") of PMS, pointing towards various classic work including Bondage by Tanaka [12]. Bondage applies direct and systematic mapping techniques to musically present photographic images, using audification and spectral filtering. These techniques are further examined in SPE in the later discussion.

The structural mapping stage in SPE, discussed in detail in the following section, could be considered as a simplified form of PMS, reduced to estimating the dynamic range and time resolution of the input data for an informed selection of the mapping target for synthetic or symbolic parameters in music. Certain mapping of the structural and error components make more sense in the design of spectral parameter encoding, while others may be useful only for human perception.

MBS fundamentally differs from both PMS and SPE with the use of interactive physical models that follow expressive natural acoustic phenomena. Somewhat similar to the assumption in SPE that musical sound structure increases the (multi-dimensional) perceptibility, MBS assumes that a well-defined virtual-acoustic system, which we may experience in a real-life natural environment, enables intuitive comprehension of complex high-dimensional data structures. The data points are mapped to the configuration or the initial state of such physical models, rather than altering the sound-producing mechanisms, therefore allowing the reuse of the same model with different data sources. In contrast, since the SPE takes the dynamic range and time resolution into account for musical organization, it benefits from and requires some level of manual examinations for each new design of musically-structured sonification.

Hermann et al. have explored other sonification techniques that are relevant to the SPE framework. For example, principal curve sonification (PCS) focuses on identifying the hidden structure across multiple data dimensions [13], while SPE attempts real-time or instantaneous extraction of the structure in a single dimension. Also, as SPE exploits the parameters of magnitude-spectrum distribution such as the mean and variance, Hermann et al. have experimented with utilizing multiple frequency bands in their spectral mapping sonification (SMS) for mapping and analyzing EEG data, enabling multi-dimensional sonification with a rich timbre [14]. While this approach may provide feasible bidirectional relationships between the input data and output sound with little channel interference, it constrains the use of polyphonic timbres to isolated frequency ranges, which SPE tries to address.

Lastly, audification is a rather straight-forward technique that maps a vector of data to the time-domain amplitude of audio samples (after some preprocessing such as scaling and filtering, if necessary). It does not involve any mapping of data structure to hierarchical representation in sound, although it may potentially reveal the structural pattern of the data as a temporal-spectral effect. The resulting audio may arguably provide the highest reversibility to the original data, given that we have access to the digital audification data or be able to capture the acoustic signal with high temporal precision with no acoustic interference. In addition to audification, various techniques exist for digitally encoding or embedding non-musical information into digital audio data, particularly in the field of audio steganography [15]. Our framework, in order to focus on perception, aims for encoding data into an acoustic signal that transmits through the air, and decoded by either a human or machine listener, we exclude discussions about purely-digital data encoding and decoding.

The following sections discuss the main components of the SPE framework with the focus on analysis-driven decision making and spectrally-decodable data mapping. First, we present the overview and the configuration of the design process. We then discuss several approaches to aligning musical and non-musical data structures by means of simple analytics. The following section elaborates the strategies and various musical techniques for designing a "reversible" sonification utilizing parameterized magnitude-spectral distribution.

2. THE FRAMEWORK OVERVIEW

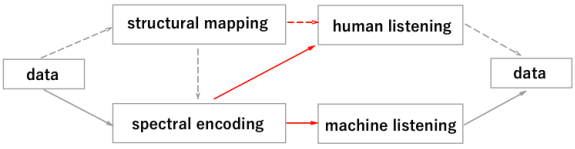


Figure 1: The basic overview of the SPE framework. The solid lines are the essential signal paths for data encoding and decoding while the dotted ones are optional perceptual treatments. Also, the red color indicates an acoustic signal as opposed to a digital signal with the gray color.

The framework for reversible musical data encoding consists of five steps: data input, structural analysis and selection of musical dimension or techniques, spectral encoding, auditory output, and evaluation & data recovery. As shown in Figure 1, some paths are optional and the whole process could consist of only the input, spectral encoding, audio output and evaluation. The structural analysis and mapping may provide additional musical organization to the spectral encoding

process. Both the mapping and encoding stages take techniques of dividing the signal into two distinct components, with somewhat different implications in each stage.

The provided examples in the following discussion are available online [1] for listening and experimenting. The example sonifications utilize a web-browser-based real-time audio environment called Data-to-Music API, developed by the authors [16], which enables various modes of synthesis and sequencing including spectral processing. As the evaluation (or the machine-listening) system is meant to be physically separate from the encoding / sonification system, the online examples do not compute the evaluation results internally. However, the reader may be able to test the examples using common audio descriptor tools, such as the zsa.descriptor library for Max/MSP [17].

3. STRUCTURAL MAPPING

As stated previously, we are interested in analyzing and repurposing the data structure for musical organization for increased expressivity and perceptibility. By "data structure", we signify not the format of the data organization but the underlying characteristics such as predictable development over time, periodicity, and distribution. SPE takes the simple approach of extracting structures using the additive error model, a commonly employed modeling technique in, for example, data compression, audio and vocal synthesis, and statistical signal processing. The analysis process separates the input signal (data) into rough structural and residual components, such that

$$x(t) = f(t) + \mathcal{E}, \tag{1}$$

where x is the input signal, f is the structural component, t is the time index, and $\mathcal{E}$ is the non-structured component. In a data-compression or audio-synthesis context, the aim of the structural decomposition would usually be the reduction of complex data into more concise parametric representations for further coding or transformation, while retaining the residual part for every data point but in a narrower and more stationary dynamic range than the original form. For the purpose of mapping to musical structures, instead of size reduction, the decomposition of data allows us a flexible mapping of non-musical data to appropriate musical structures without losing the information as a whole.

3.1. Examples of Structural Mapping

Figure 2. An example of decomposition of structure (Left: Input, Center: Estimated structure, Right: Residual signal)

To illustrate the structural mapping process in a very simple sonification problem, suppose we encounter a single-dimensional time series with slowly increasing central values with somewhat stationary noise deviating around them (Figure 2), which may be expressed as

$$x(t) = \frac{1}{1 + e^{-t}} + \varepsilon. \tag{2}$$

Mapping the original signal x to, for example, the volume of an oscillator (Online Example 1) has a potential of impeding the musical balance or perceptibility as the slow increase of the volume may be hard to hear in the beginning, and it does not take advantage of the dynamic range of human hearing at any given moment. Similarly, if we map x to the frequency of an oscillator with a fixed amplitude (Online Example 2), although it may represent the data faithfully, it may also produce a sense of "unstable" pitch slowly evolving that does not reside well in more complex sonifications where multiple sound dimensions are presented. The extraction of a larger non-stationary envelope allows repurposing of such non-musical data sources by, for instance, mapping the residue to a full-range of amplitude to hear the detailed fluctuations while the slow-moving central values could be assigned to the frequency to produce more stable and "organized" pitch-sweeping gesture (Online Example 3).

With unordered data, the focus of analysis typically shifts to, for example, clustering, cross correlation, or observing the distribution. In SPE, we utilize the shapes of distribution for musical organization. For example, if a given distribution does not fit to a Gaussian function, we may instead parameterize it into line segments with a fewer number of breakpoints. We can then generate a percussive or metallic timbre with sinusoidal oscillators with the frequency randomly sampled from the parameterized distribution, as described in the next section, while using the residual signal for amplitude modulation (Online Example 4).

Another approach for utilizing the value distribution rather imposes an existing musical structure, such as a musical scale that introduces the minimum amount of distortion, to transform the data non-linearly while retaining the residual values for additional parameter encoding. The Online Example 5 demonstrates the selection of the best musical scale by computing the signal-to-noise ratio after quantizing the data points with the common musical scales in all transpositions. This example may provide improved musical perceptibility, but does not assure a spectrum-based data recovery. For that, the quantization residual signal from may be mapped to, for example, a parameter of the magnitude spectrum, as discussed in section 4.1.2.

3.2. Estimation of Data Structure

To roughly estimate a non-stationary envelope structure in unknown data, we may naively apply a high-order moving-average filter (mocking iterative linear prediction analysis) and then examine the uniformity of the residual noise by calculating the normalized entropy (or variance) of the distribution (Online Example 6). The resulting residual signal is not necessarily uncorrelated (i.e., may contain periodic patterns), but the increase of stationary quality enables more optimal mapping to certain musical parameters. In addition to finding a larger structure, a spectral filter may also capture high-frequency repetitions (e.g., by isolating the smaller coefficients in discrete cosine transform) (Online Example 7). This approach is similar to separating the salient resonances and the noise floor in spectral modeling synthesis [18], in which the noise floor can be replaced with a parameterized spectral noise generator. In addition to these contour

[1] https://GTCMT.github.io/DataToMusicAPI/icad2017
(Online Examples, Accessed May 15, 2017)

extractions, predictive analysis or the first-order differential of the signal may capture local sequential dependencies in the data.

Although more complex and multi-dimensional statistical analysis using iterative computation may be beneficial, our motivation of using a relatively naive estimation is to allow real-time design (e.g., live coding [19]) of sonification with unseen data, including short-time analysis and mapping of streaming data. The simplicity of the analytical process also helps in retaining intuitive relationships between the input signal and the output parameters, which take relatively close-to-original data structure, while dynamic transformations such as dimensionality reduction may not be suitable for the purpose of the data recovery from the resulting audio.

Therefore, the use of an additive model is beneficial in that it provides a coarse structural component for human perception (and potentially for MIR classification tasks or a model-based estimation of a cleaner signal). On the other hand, the residual component is suited for preserving fine details that may be deterministically recovered by spectral feature extractors [6][17]. Combining the structural and residual parts, after proper rescaling, produces a close-to-original estimate of the input data.

4. SPECTRAL PARAMETER ENCODING

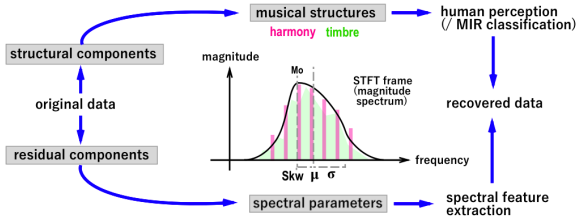


Figure 3. A signal flow incorporating structural analysis (Section 3) into spectral parameter encoding.

The framework aims for not only aligning the musical dynamics for multi-dimensional comprehension, but also preserving as much information as possible in the acoustic signal for computational data retrieval. As shown in Figure 1, this spectral encoding process may be applied directly and entirely to the raw incoming data, or may be combined with a structural analysis to utilize extra information in the mapping and generation processes. A sensible approach may be to route the structural element of the data to the selection of musical expressions to generate, such as harmony and timbre, while using the subtracted residual part for spectral parameters that define the acoustic contour (see Figure 3). However, both signal paths may also be used for spectral encoding, while the musical content is kept as ornamental or non-essential for computational data recovery.

Similar to the additive model used in the structural mapping, the spectral parameter encoding combines two parts, statistical magnitude-frequency parameters and a matching distribution, to generate an acoustic result, such that:

$$y = IFFT[g(\theta_t, X)]. \quad (3)$$

The distribution parameter variables, θ, deterministically holds the input data (that are linearly scaled if necessary), which are later measured by spectral descriptors to estimate the original (or post-scaled) values. The encoded parameters may be the first-order descriptive statistics such as spectral centroid (weighted mean), spectral spread (variance), skewness (median), and spectral crest factor (tonality measure). The actual contents of the spectral distribution do not matter as long as they satisfy the statistical analysis in the short-time Fourier transform (STFT) signal [20]. This allows us to employ various approaches for creating musical expressions, including timbral, harmonic (with musical scales), and polyphonic mixed-timbre voices, that are expanded from the target parameter values.

4.1. Spectral Encoding Techniques

Here, we discuss several techniques of musical sound composition that conform to the acoustic constraints for a sufficient level of data retrieval. The encoding part may utilize either time-domain or frequency-domain synthesis via discrete-time Fourier transform (FT), while the analysis takes part in the FT of the output audio signal.

4.1.1. Timbral Structures

First, we present several of the timbre-based expressions. After specifying the distribution function parameters such as the mean and variance from the input data, one may use spectral filtering to create percussive or sweeping ambiance with a changing noise-color (Online Example 8), similar to Tanaka's Bondage discussed previously. This encoding technique requires a frequency-domain element-wise multiplication of the given envelope to the STFT of a white noise, then inverting to the time-domain signal. The resulting audio enables a relatively robust data retrieval with spectral feature description (e.g., centroid and spread) or envelope estimation. Aside from amplitude changes over time such as attack and decay shapes, the spectral content (white noise) may not provide additional structure for perceptual sonification compared to other techniques discussed below. However, the distribution function may take an elaborate shape with linearly-interpolated break points, which would be analyzed with peak estimation or band-limited magnitude analysis similar to SMS. While this approach is efficient for real-time synthesis, it requires the matching of the input vector length and a half of the FT frame by, for example, linear interpolation or zero-padding the extreme frequency ranges.

Though the spectral filtering approach may be good for creating expressions of generic noise percussion, its timbral dynamics can be fairly limited. Instead of using a time-domain noise generator, the oscillator-bank-based approach allows us to take the same spectral distribution function but create more focused (pitched) timbre with non-harmonic partials suited for metallic percussion or ambient pad sounds. This may be realized with a granular (i.e., random-phase, Online Example 9) or an additive (i.e., synchronized-phase, Online Example 10) synthesis using, for example, the random-sampling technique using the inverse transform [21] of the cumulative spectral distribution.

The additive synthesis in the oscillator-bank approach is similarly robust for the recovery of data as spectral filtering. However, the granular synthesis tends to introduce phasing interferences among partials, making the estimation of the statistical parameters less reliable. Since the random time-domain source signal in these timbral-composition approaches cannot be easily estimated, they may be suited for representing the residual noise envelope from the structural analysis step. The structured component, such as a slow-moving contour, may be utilized as gain envelope after normalizing the magnitude spectrum.

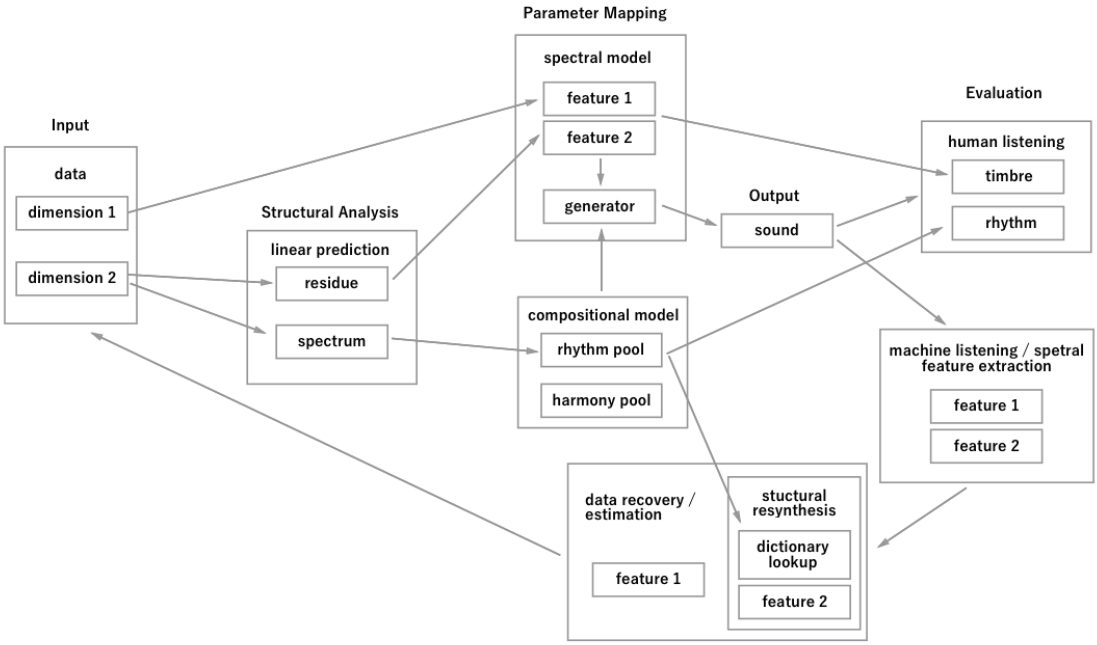


Figure 4. An example of multi-dimensional SPE system

4.1.2. Harmonic Structures

In a more symbolic-level musical organization, one can take spectral parameters, in which we encode data, and generate a single note with natural harmonics or multiple notes forming an arbitrary harmony. For instance, given a spectral centroid (weighted mean), it is trivial to expand it to a single note with N harmonic partials with a fixed unit amplitude by

$$\begin{aligned} n &= 1, 2, \dots, N;\ N \in \mathbb{Z}, \\ v_n &= \frac{nN\mu}{N!}, \end{aligned} \tag{4}$$

where v is the vector of frequencies for sinusoidal additive synthesis, μ is the spectral centroid in Hz, and N is the number of the harmonics (Online Example 11). We can also generate any pitch by adjusting the amplitude and number of overtones and undertones accordingly. The following example computes a single tone conforming to given spectral centroid and spectral spread (weighted variance) values. For an odd number of natural harmonics with an identical gain for non-central oscillators,

$$\begin{aligned} &for\ N\ \in \{1, 3, 5, \dots\}, \\ &n = 1, 2, 3, \dots, N, \\ &a_n = \begin{cases} 1\ when\ n = \dfrac{N+1}{2} \\ g\ otherwise \end{cases} \\ &g = \frac{\sigma^2}{\sum (v_n - \mu)^2 + \sigma^2 (1 - N)}, \end{aligned} \tag{5}$$

where a is the gain coefficients with a symmetric form $\{\dots, g, g, 1, g, g, \dots\}$ and σ is the square root of the spectral spread (i.e., standard deviation, Online Example 12). Similarly, we can construct an arbitrary harmony with sinusoidal oscillators centered around a given spectral centroid, such that

$$v_n = \sum_{n=0}^{N-1} \frac{\mu C_n}{\bar{C}}, \tag{6}$$

where C is a vector of normalized frequency coefficients in Hz for creating a chord[2] and $\bar{C}$ is the average frequency of them (Online Example 13). Combining both additive synthesis and harmony, it is also feasible to generate any chord on any root note from given spectral parameters (Online Example 14). The flexibility in generating the pitch or the chord quality can be utilized to encode additional data dimensions for human perception. These harmonic techniques are relatively robust for retrieving the spectral parameters, provided that the data mapped to the spectral centroid is scaled properly so that the harmonic or chord-voice frequencies exist within the spread range.

4.1.3. Further Applications

Lastly, in addition to encoding data to multiple spectral parameters for generating a single spectral distribution, the potential of SPE is the ability to mix multiple timbral techniques (e.g., harmony model + single-additive-voice + noise percussion) as long as they conform to the overall distribution parameters (Online Example 15), or even to mix multiple distributions and estimate their parameters with a mixture-model parameter estimation [22]. Mixing multiple instruments to a single distribution enables additional

[2] This may be obtained by converting a list of degrees in MIDI note number starting at 0 (e.g., [0, 4, 7, 11] for the major seventh) to frequency.

perceptual dimensions that may be tracked by the human listener (e.g., the salience of a particular instrument), while preserving the most critical data channels in the spectral parameters for computational recovery.

5. DISCUSSION AND FUTURE WORK

To summarize, SPE encapsulates the data points, either from raw input or analyzed structures, to the abstract statistical shape or parameters (e.g., mean and variance) of a magnitude-spectrum frame. This facilitates a uniquely constrained yet flexible composition expanded from the target magnitude spectrum, and even allows additional mapping of data to such as the choice of chord or onset shape for perceptual decoding. As we include multiple data dimensions in the analysis, mapping, and encoding paths, the entire signal flow may grow into a quite complex system as Figure 4.

We did not, however, examine musical expressions that extend over several seconds (e.g., rhythms, melodic patterns) in this discussion. For future work, we plan on examining spectral encoding techniques over time utilizing symbolic parameterizations. Relevant work includes Smalley's spectromorphology [23], an analytical framework for electro-acoustic music in which the author lists qualitative distinctions in each morphing (moving) steps of spectral contents. The chosen parameters (e.g., "upbeat" + "transition" + "closure") combine and form a complex musical gesture over time. In addition, spectral modeling synthesis [18] also provides insights in creating time-varying timbral structure with the deterministic and random components.

The time-varying encoding poses a practical issue with the time resolution of the data stream. SPE analyzes the STFT frames of the output audio with a reasonable frequency resolution (e.g., 1024 samples at the sampling rate of 44100 for harmonic or granular approach), which limits the data rate to at least one datum per 20 milliseconds. This is quite slow compared to, for example, audification or even possibly PMS. The data rate is forced to decrease even more when using encoding techniques such as granular synthesis or mixed-timbre composition because of the susceptibility to voice phasing. Also, adding time-domain audio effects such as delay and reverberation also smears out the phase relationship, causing more errors in machine listening.

6. CONCLUSION

We presented spectral parameter encoding, a dual-layer framework for musically expressive yet functional design of sonification. It employs a simple structural analysis to facilitate a semi-automated organization of mapping, and data encoding to spectral features as well as computational feature extraction to ensure the minimized loss of information as a whole in the process of transformation and mapping. Although the use of spectral distribution imposes certain acoustic constraints, it allows a variety of musically-organized sonification from timbral to harmonic expressions with the possibility of a multi-timbral structure.

7. REFERENCES

[1] P. Vickers and B. Hogg, "Sonification Abstraite/Sonification Concrete: An 'Aesthetic Persepctive Space' for Classifying Auditory Displays in the Ars Musica Domain," in *Proceedings of the 12th International Conference on Auditory Display (ICAD 2006)*, 2006.

[2] C. Roads, *Microsound*. Cambridge: MIT Press, 2001.

[3] N. Degara, F. Nagel, and T. Hermann, "Sonex: An Evaluation Exchange Framework For Reproducible Sonification," Jul. 2013.

[4] K. Vogt, "A Quantitative Evaluation Approach to Sonifications," Jun. 2011.

[5] A. Klapuri and M. Davy, *Signal Processing Methods for Music Transcription*. New York: Springer, 2006.

[6] A. Lerch, *Audio Content Analysis: An Introduction*. Hoboken, N.J.: Wiley, 2012.

[7] B. Logan, "Mel Frequency Cepstral Coefficients for Music Modeling."

[8] F. Grond and J. Berger, "Parameter Mapping Sonification," *Sonification Handb.*, pp. 363–397, 2011.

[9] T. Hermann and H. Ritter, "Listen to your Data: Model-Based Sonification for Data Analysis," in *189–194, Int. Inst. for Advanced Studies in System research and cybernetics*, 1999, pp. 189–194.

[10] T. Hermann, "Model-Based Sonification," *Sonification Handb.*, pp. 399–427, 2011.

[11] R. L. Alexander, J. A. Gilbert, E. Landi, M. Simoni, T. H. Zurbuchen, and D. A. Roberts, "Audification as a Diagnostic Tool for Exploratory Heliospheric Data Analysis," Jun. 2011.

[12] A. Tanaka, "The Sound of Photographic Image," *AI Soc.*, vol. 27, no. 2, pp. 315–318, May 2012.

[13] T. Hermann, P. Meinicke, and H. Ritter, "Principal Curve Sonification," 2000.

[14] T. Hermann, P. Meinicke, H. Bekel, H. Ritter, H. M. Müller, and S. Weiss, "Sonification for EEG Data Analysis," in *Proceedings of the 2002 International Conference on Auditory Display*, 2002.

[15] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, "Comparative Study of Digital Audio Steganography Techniques," *EURASIP J. Audio Speech Music Process.*, vol. 2012, no. 1, p. 25, 2012.

[16] T. Tsuchiya, J. Freeman, and L. W. Lerner, "Data-to-Music API: Real-Time Data-Agnostic Sonification with Musical Structure Models," *Proc 21st Int Conf Audit. Disp.*, 2015.

[17] M. Malt and E. Jourdan, "Zsa.Descriptors: a library for real-time descriptors analysis," *ResearchGate*.

[18] X. Serra and J. Smith, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition," *Comput. Music J.*, vol. 14, no. 4, pp. 12–24, 1990.

[19] T. Tsuchiya, J. Freeman, and L. W. Lerner, "Data-Driven Live Coding with DataToMusic API," in *Proceedings of the 2nd Web Audio Conference (WAC-2016), Atlanta*, 2016.

[20] R. Bracewell, *The Fourier Transform and Its Applications*, vol. 5. 1965.

[21] S. Olver and A. Townsend, "Fast Inverse Transform Sampling in One and Two Dimensions," *ArXiv13071223 Math Stat*, Jul. 2013.

[22] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digit. Signal Process.*, vol. 10, no. 1, pp. 19–41, Jan. 2000.

[23] D. Smalley, "Spectromorphology: Explaining Sound-Shapes," *Organised Sound*, vol. 2, no. 2, pp. 107–126, Aug. 1997.

DID YOU FEEL THAT? DEVELOPING NOVEL MULTIMODAL ALARMS FOR HIGH CONSEQUENCE CLINICAL ENVIRONMENTS

Parisa Alirezaee, Roger Girgis*, TaeYong Kim*, Joseph J. Schlesinger†, and Jeremy R. Cooperstock**

Centre for Interdisciplinary Research in Music, Media and Technology, Montreal, Canada
* Department of Electrical and Computer Engineering, McGill University, Montreal, Canada
† Depts. of Anesthesiology & Biomedical Eng., Vanderbilt University Medical Center, Nashville, USA
{parisa|rogerg|taeyong|jer}@cim.mcgill.ca
joseph.j.schlesinger@vanderbilt.edu

ABSTRACT

Hospitals are overwhelmingly filled with sounds produced by alarms and patient monitoring devices. Consequently, these sounds create a fatiguing and stressful environment for both patients and clinicians. As an attempt to attenuate the auditory sensory overload, we propose the use of a multimodal alarm system in operating rooms and intensive care units. Specifically, the system would utilize multisensory integration of the haptic and auditory channels. We hypothesize that combining these two channels in a synchronized fashion, the auditory threshold of perception of participants will be lowered, thus allowing for an overall reduction of volume in hospitals. The results obtained from pilot testing support this hypothesis. We conclude that further investigation of this method can prove useful in reducing the sound exposure level in hospitals as well as personalizing the perception and type of the alarm for clinicians.

1. INTRODUCTION

The operating room (OR) and intensive care unit (ICU) are noisy environments, exacerbated by frequent auditory alarms. As an example, the average sound level of patient rooms in three Veterans Affairs facilities were measured at 51 dB(A) [1], whereas the World Health Organization (WHO) recommends sound levels 15-20 dB(A) lower to avoid negative impact on patients and staff, such as sleep deprivation and alarm fatigue.

The high incidence of alarms in the ICU and OR command attention, raise stress, and are often irrelevant to the responsibilities of individual clinicians. In multi-bed care areas in hospitals, one can count more than 30 different alarm sounds [2]. The increasing number of alarms in hospitals cause problems because of the lack of clinical information provided by their interfaces and the stressful environment caused by their overall volume. Although nurses and doctors rely on patient-monitoring devices for diagnosis and treatment, a high number of these devices increase the rate of false alarms by not reflecting a medically urgent condition. This, in turn, may lead to error-prone situations.

https://doi.org/10.21785/icad2017.066

Additionally, loudness of the alarms can cause "alarm fatigue", which is a phenomenon of diminished response due to desensitization of the practitioners [3]. This problem is exacerbated by the fact that free-field audible alarms are not informative, non-localizable, and presented to everyone in the room. When the fatigue is severe, it can increase clinician error, potentially leading to adverse patient outcomes.

To cope with these problems, we are interested in the possibility of leveraging multimodality to improve the information transfer capacity of alarms. Through this approach to alarm management, we anticipate a reduction in overall sound exposure level in the clinical environment and decreased distractions, as well as a possibility of implementing personalized alarms.

An extensive volume of previous literature has described the effects of multisensory integration, in particular involving the auditory and visual modalities. These include evidence for both complementary and inhibitory effects of the combination [4]. However, visual attention is directional, and in the case of a clinical environment, one cannot assume that a visual signal would be attended to by the health-care provider. In contrast, haptic feedback can be provided anywhere, any time, irrespective of current activity, and offers the additional benefit that it can be delivered selectively to the clinician(s) for whom it is relevant.

To investigate the possibility of leveraging the audio-haptic modality in this manner, we conducted an experiment to determine the degree to which haptic stimuli can complement audible alarms. To quantify the accuracy to such cues, we compared unisensory auditory and multisensory auditory-haptic stimuli.

The results of our experiments did not demonstrate a facilitative effect as expected. However, analyzing them raised important questions regarding fatigue and habituation to vibrotactile stimuli, potential interference of sensory streams, potential benefits of speech over non-speech auditory stimuli, and if supra-threshold stimuli can still be weakly effective as a unisensory stream to contribute to multisensory gain.

2. LITERATURE REVIEW

To understand the importance of studying alarm systems in the hospital setting, Block divides the set of problems with audible alarms into several categories: 1) false alarms; 2) loud alarms; 3) difficulty in determining the device that is making the alarm sound; and 4) inability to stop an alarm [5].

These problems worsen with the fact that devices designed by different manufacturers may use different alarm sounds to indicate the same event occurrence. Block then discusses how an alarm sound can address a specific medical situation. His approach grouped audible alarms into a set of six different sounds, each related to one tissue injury situation for the patient, and assigned a different priority. He tested different melodies for these alarm sounds as well as the training method to instruct the clinicians to determine the sounds and their meanings.

In a similar vein, Edworthy focuses on false alarm events in hospitals [3] [6], investigating important issues related to the design, implementation, classification, and standardization of medical alarms. Although the design of medical alarms may seem trivial, the literature makes clear that despite the large investment in the development of standards, i.e., IEC 60601-1-8 [7], significant problems remain. For example, Talley et al. found a false alarm rate of 85% to 99% in cardiopulmonary monitors [8]. Unfortunately, this level of false alarms has not changed significantly over the past 25 years. As another approach Keller tackles the false alarm problem by reducing more generally the number of alarms step-by-step to reach an acceptable level [9]. For each step, a safety assessment analyzed the response process to a task on alarm notification. This approach succeeded in reducing the number of alarms from a high number to around four per day.

The aforementioned issues are recognized by regulatory bodies. In the 2015 top 10 Health Technology Hazard list published by the Emergency Care Research Institute (ECRI), the top priority was stated to be clinical alarm hazards. Nurses spend an unnecessary amount of time on alarm management and on dealing with ambiguous alarm implementations. According to the U.S. Food and Drug Administration (FDA), in the span of four years, more than 500 patient deaths have been associated with faulty alarm management [10]. Despite these issues, audible alarms tend to be the most effective method in attracting the attention of practitioners when they are occupied with other tasks.

However, as audible alarms became the prominent method, their use had the unintended consequence of Alarm Fatigue [10]. Without a formal operational definition, this phenomenon can be understood as the desensitization to alarm sounds as a result of hearing too many alarms at the same time. Thus, the medical staff is incapable of responding to alarms with equal urgency.

With multiple alarm stimuli, there is attentional competition and auditory competition (e.g., masking). Hasanain et al. explored the issue of simultaneous masking; a condition where because of the interaction between concurrent sounds, one or more of them become imperceptible due to the physical limitation of human perception [11]. According to the Pennsylvania Patient Safety Authority, 194 documented problems with operators responses to alarms over a span of four years resulted in twelve deaths [12]. The cause was found to be related to the number of simultaneous alarms sounding in the medical environment. Detecting alarm masking is highly challenging as it may occur only in a very specific interaction of multiple sounds. Hasanain et al. proposed a method to do so without experimentation [11]. Instead, the authors used psychophysical modeling in the configuration of medical alarms, which builds on previous work done by Hasanain and Bolton [13].

The problem of "Alarm Fatigue" has been investigated by understanding the psychoacoustic properties of alarms and analyzing the sensory perception of clinicians. Ongoing experiments by Schlesinger et al. [14] are designed to determine the auditory perceptual threshold of alarms while participants perform other attentional demanding audiovisual tasks. Hospital alarms are typically louder than background noise, i.e., a positive signal-to-noise ratio. However, their results showed that clinician performance measured in response time and accuracy was preserved when alarms were softer than background noise, i.e., a negative signal-to-noise ratio.

The method we propose in the present work attempts to employ multisensory integration; specifically, in combining the audio and haptic channels, with the goal of reducing the sound exposure level, and therefore stress level, in the hospital environment. The Principle of Inverse Effectiveness (PoIE) [15] suggests that an enhanced neural response can be achieved when stimuli from two modalities are simultaneously presented. This effect becomes greater if the stimuli produce a weak response when presented unimodally. Experiments have shown this perceptual additivity at sub-threshold levels from neural inputs of the olfactory and gustatory channels [16]. Furthermore, co-occurrence of a sound was found to increase accuracy and enhance the sensitivity for detection of near-threshold visual stimuli [17].

The PoIE led us to theorize that the co-occurrence of a haptic stimulus and an audio stimulus would allow participants to perceive sound below their audio threshold of perception. In the OR and ICU setting, this would allow a reduction of the alarm intensity while maintaining the effectiveness in cueing a staff member. Furthermore, it may be advantageous to present the haptic stimulus at a sub-threshold level to not disrupt the procedure being performed by the prompted medical professional. Visell et al. [18] showed that the addition of sub-threshold stimuli affected the participants perception of compliance. This provides encouraging evidence that even at a sub-threshold level, such stimuli may be sufficient to affect alarm perception.

3. METHODOLOGY

There is evidence from the literature to support the hypothesis that multisensory integration may lead to participants perceiving sound at a lower threshold. However, we are equally interested in determining whether this effect may hold when the non-audio stimulus is delivered at a sub-threshold level. That is, can we reduce the level of auditory alarms in a clinical environment by delivering a complementary non-auditory stimulus, ideally, one that the clinician does not even perceive?

In order to investigate this question, it was first necessary to determine the unimodal thresholds of perception for both the auditory and non-auditory stimuli. Our experiment therefore consists of three measurements:

1. haptic (vibration) perception threshold
2. auditory perception threshold
3. auditory perception threshold when combined with haptic stimulus

One of the most popular methods to map the relationship between physical stimuli and psychological response of the participant [19] is Parameter Estimation by Sequential Testing (PEST) [20], [21], an adaptive staircase method that has shown its adaptability and robustness in obtaining a perceptual threshold value. Figure 1 represents a typical double staircase for measuring the auditory threshold for one of the participants. An improvement is to double the step size in response to several identical responses,

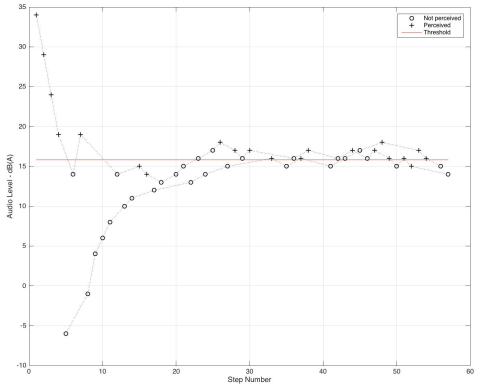


Figure 1: An example of the double-staircase method to determine the auditory threshold of perception for one of the experiment participants.

and halve the step size in response to a change in consecutive responses. This helps achieve faster convergence and improves participant focus, and was therefore adopted for our testing [19].

To reduce the effects of bias that arises after several identical responses to a given stimulus, Cornsweet suggested the use of the random double-staircase method [22]. The test participant is presented with two staircases, starting from values above and below the assumed threshold, respectively. The step size begins relatively high to ensure fast convergence, and as the two staircases approach each other, the step size is reduced to ensure a smooth combination of the two. This results in a range of values bounding the threshold of perception.

Throughout our experiment, we employ the PEST procedure coupled with the use of the random double staircase to determine the threshold of perception.

4. EXPERIMENTS

4.1. Environment

The experiment was performed in the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). The lab is acoustically insulated from the surrounding rooms. In addition, participants wore Beyerdynamic DT 770 Pro (Heilbronn, Germany) circumaural headphones during the experiment, and the ambient temperature was maintained between 21 and 25 throughout the tests, thereby ensuring a well-controlled environment.

4.2. Stimuli

To provide the vibrotactile stimulus, we used a Tactile Labs Haptuator Mark I (Montreal, Canada) [23], which allows for independent variation of the amplitude and the frequency. For our experiment, the haptuator was connected to a Sparkfun TP2005D1 audio amplifier (Boulder, CO, USA), and strapped snugly to the participants' leg, above the ankle, using a Velcro band. The choice of placing the vibrational device on the leg rather than the wrist was motivated by our intended use case of delivering alarm signals in a medical environment, for which hygienic constraints preclude the wearing of devices on the hands or wrists. A potential confound is that actuation of the vibrational device may be audible, contributing to the sound volume of the alarm stimulus. However, this was mitigated by the low intensity of vibration, the placement of the actuator on the participants' leg, and the use of closed headphones throughout the experiment.

Delivery of stimulus during the experiment and logging of measurements for the double staircase was managed by a MATLAB script (MathWorks MATLAB R2016a, Natick, MA, USA). A one-second auditory stimulus was extracted from a recording of the Philips MP-70 (Amsterdam, Netherlands) patient monitor red/crisis alarm. The frequency spectrum of the alarm sound is shown in Figure 2. The choice of a one-second duration was deemed to be reasonably short to help eliminate guesses, and sufficiently long so as to include the salient auditory characteristics of the alarm signal.

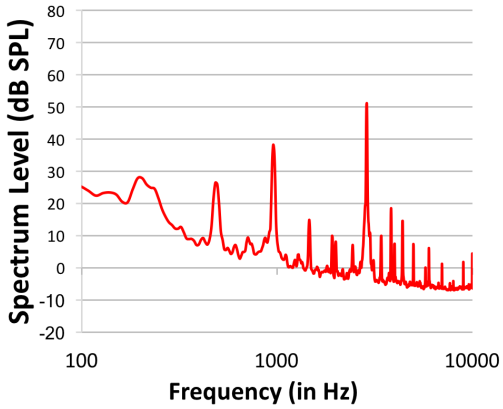


Figure 2: Alarm waveform, measured with a class II Amprobe SM20A sound level meter provided a weighted output of 49 dB.

The vibratory stimulus was generated using a sine wave at 175 Hz, output using the MATLAB sound() function for a duration of 1 second. For the combined auditory-haptic stimuli, the signals were output in unison, using a stereo audio splitter to separate the audio (left channel) and haptic (right channel) signals.

4.3. Experimental Procedure

Participants first completed a pretest questionnaire to screen for possible health conditions that might exclude them from the experiment. They were then asked to read an instruction sheet, put on the headphones, and assisted with securing of the haptic band just above the ankle at a comfortable location, as shown in Figure 3.

The strap was secured in a snug fashion for good coupling between the actuator and the skin, but not so tight that it caused discomfort. Participants were then asked to place their right foot in a comfortable position and to immobilize it for the remainder of the experiment.

Participants then proceeded through the first block of the experiment, which determined their haptic perceptual threshold. In the second block, we determined the auditory perceptual threshold, both with and without a combined haptic stimulus. This was done by intermingling two tests, randomly selecting half of the trials for presentation of unimodal auditory stimuli, and half for presentation of combined audio-haptic stimuli with the haptic stimuli

Figure 3: Position placement of the actuator on the participants' ankle, from experiment instructions.

delivered at a fixed level. This intermingling was done to avoid potential habituation effects that may have biased the threshold estimates in either direction.

For each experimental condition, participants initially carried out a training/calibration step to familiarize themselves with the experimental stimuli and adjust these to a level in which they were barely perceptible, using a coarse staircase method. Subsequent to this initial adjustment, the PEST method, using a dual-staircase procedure, was employed to determine the participants' perceptual thresholds. Each stimulus was delivered randomly within an 8 s window following the previous stimulus.

To ensure that participants did not go too long between successive perceivable stimuli, which we observed during pilot testing as a significant source of fatigue, the system randomly delivered 20% of the stimuli at 3.5 standard deviations above the average of the intensities of the last six stimuli, a choice we determined empirically as adequate to ensure reasonable supra-threshold perception.

For our purposes, participants had to respond within 2 s from the onset of the stimulus, i.e., no more than 1 s following its presentation, by clicking on a button displayed in a simple graphical user interface; otherwise, it was assumed that they did not perceive the stimulus. The system then reduced or increased the subsequent stimulus intensity by a defined step size so as to maintain the intensity just at the edge of perceptibility. The step size was reduced as both staircases converged, i.e., as the difference in intensities between the upward and downward staircases decreased. This process continued until a minimum step size was reached, then six reversals were counted on each staircase for the threshold estimation. The perceptual threshold was estimated as the mean of the stimulus levels of the last six reversals from each staircase.

The aforementioned design was employed for all of our experiments, with variations as described in the following sections.

4.4. Pilot

Initially, we fixed the sub-threshold amplitude of vibration at 2.5 standard deviations below the perceptual threshold determined in the first block.

Pilot testing was performed on 11 lab members (10 male, 1 female), over the span of three days. The mean age of the participants was 30.4 years with a standard deviation of 8.6 years. Data from one of these participants were excluded from the analysis for failing to respond to supra-threshold "wake-up stimuli", which suggested a lack of attention during the experiment. The test participants participated on a voluntary basis and did not receive monetary compensation for their time.

Five of the 10 participants whose data were retained for analysis exhibited a slightly lower auditory threshold when measured in the multimodal audio-haptic condition than in the unimodal audio-only condition. Although the results of these initial tests were only borderline in terms of statistical significance ($p = 0.15$, ci = [-0.74, 0.13]), we were encouraged to carry out a larger experiment with additional participants who were naive to the experimental hypothesis.

4.5. Full Experiment

The same experimental procedure was then applied to a new group of participants, naive as to the experimental hypothesis. These participants were not informed that sub-threshold vibration was delivered (in conjunction with half of the audio stimuli) during the second block. The test was conducted on 12 participants (10 male, 2 female), over the span of three days. The mean age of the participants was 28 years with standard deviation of 3.5 years.

Data from one of these participants, exhibiting a difference between the audio-only and audio-haptic thresholds greater than six times, were excluded from the analysis as an outlier. The test participants participated on a voluntary basis and received monetary compensation for their time. The auditory threshold values measured across 12 participants for both the audio-only condition and the audio-haptic condition is shown in Figure 4. The differences between the conditions are shown in the bar graph of Figure 5.

As can be seen, the data did not support our hypothesis that the perceptual threshold is reduced in the multimodal condition. Counter-intuitively, the trend suggested an opposite effect, although not significant ($p = 0.78$, ci = [-0.65, 0.83]). This led us to consider the possibility that our haptic stimuli was too far below the perceptual threshold, and was thus not contributing to the effect.

4.6. Increased Haptic Intensity Level

To address the possibility that the haptic stimuli were too far below threshold to have an impact, we then conducted a further experiment in which the amplitude of vibration was increased to 0.5 standard deviations below the threshold determined in block 1. As before, participants were not informed that they might feel vibrations during the third block.

Ten participants (4 male, 6 female) were recruited and the test took place during one day. The mean age of the participants was 27.3 years with standard deviation of 5.7 years. Data collection from one of the participants could not be completed on account of the participant changing the computer's output volume in the middle of the test. This participant was excluded from the analysis. The test participants participated in the study on a voluntary basis and received monetary compensation for their time.

The perception threshold in the auditory domain measured across 9 participants for both the unisensory and multisensory conditions is shown in Figure 6. The differences between the conditions are shown in the bar graph of Figure 7.

By way of response to a post-test questionnaire, 5 out of 9 participants indicated that on occasion, they perceived the haptic stimulus during the second block.

Despite the increase of the level of haptic stimulus, we found no statistically significant difference between the threshold of audio perception in the unisensory and multisensory conditions ($p = 0.21$, ci = [-0.90, 0.23]).

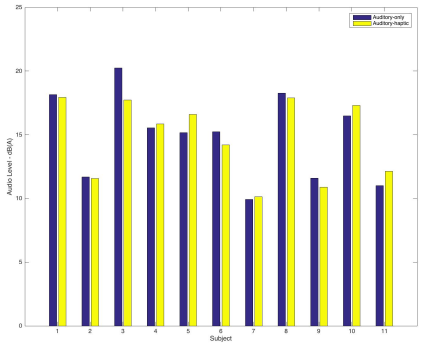


Figure 4: Threshold data obtained from the audio-haptic and audio-only threshold measurement over the 11 participants whose data were retained.

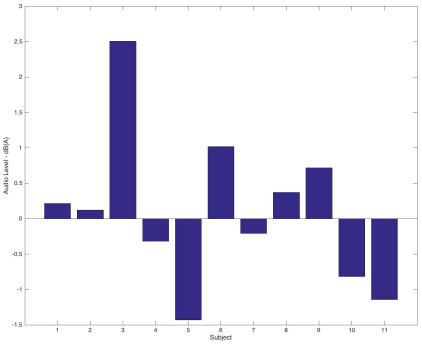


Figure 5: Differences between audio-only and audio-haptic threshold values from Figure 4. A positive value indicates that the auditory perceptual threshold was reduced in the multimodal condition.

5. CIRCLING AROUND THE THRESHOLDS

We considered several possibilities for the results of the previous section:

1. Either or both auditory and haptic perception thresholds varied throughout the experiment, e.g., due to fatigue or habituation, and so the thresholds measured in blocks 1 and 2 were unreliable.
2. The presentation of simultaneous sub- or near-threshold haptic stimuli interfered with auditory perception.
3. There is no multimodal integration benefit from haptic stimuli in conjunction with non-speech audio.
4. There is a multimodal integration benefit from haptic stimuli in conjunction with non-speech audio, but only for supra-threshold haptic stimuli.

To examine these possibilities, we conducted an additional exploratory experiment, in which we varied the level of auditory stimuli in a range of $\pm$ 2 standard deviations and varied the level of haptic stimuli in a range of [-2,+4] standard deviations around the unimodal thresholds determined in blocks 1 and 2.

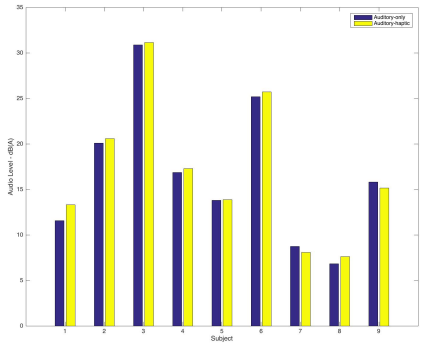


Figure 6: Threshold data obtained from the audio-haptic and audio-only threshold measurement over the 9 participants from the increased haptic intensity level experiment.

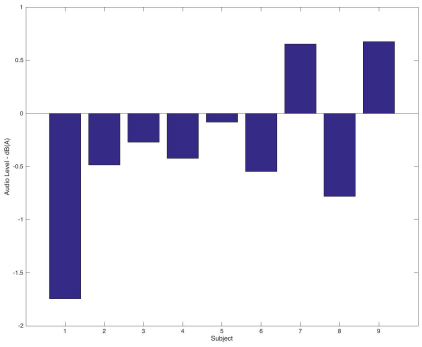


Figure 7: Differences between audio-only and audio-haptic threshold values from Figure 6. A positive value indicates that the auditory perceptual threshold was reduced in the multimodal condition.

The results of this exploration were revealing, although hardly conclusive: three of the five participants, two of whom were co-authors of this paper demonstrated no discernible effect of haptic stimulus on audio stimulus detection, even at clearly supra-threshold levels of haptic stimuli. The remaining two participants demonstrated a possible effect, with slightly higher rates of audio alarm detection for sub-threshold audio when presented in conjunction with supra-threshold haptic stimuli (see Figure 8 for the results of one of these participants). However, it does not appear that these results are significant, and thus, we can neither confirm nor reject any of the possibilities described at the start of this section.

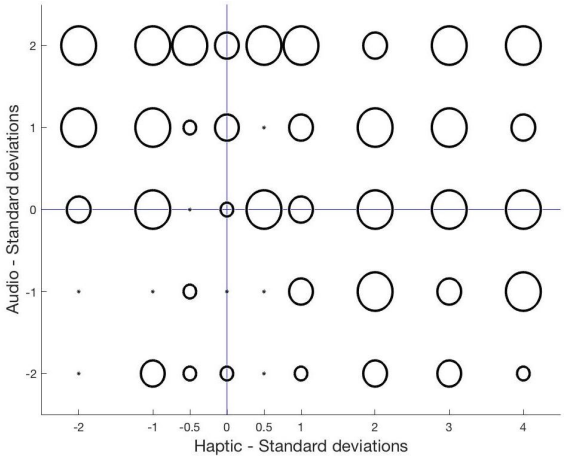


Figure 8: Detection rate for one participant of audio stimuli as a function of audio and haptic stimulus level, relative to the thresholds measured in blocks 1 and 2. The size of each circle indicates the detection rate (out of three presentations) at each combination of parameters.

6. CONCLUSIONS AND FUTURE DIRECTIONS

We postulated at the outset of this research that improved perception of auditory stimuli would result through multimodal integration with a complementary haptic signal, possibly even at subthreshold levels. If so, we hypothesized that this could allow for attenuation of alarm fatigue and assist the practitioner in recognizing the alarm, thereby reducing the problems of stress and alarm fatigue in the clinical settings of the OR and ICU.

Through these experiments, we hoped to determine preliminary guidelines for the outcomes of implementing multimodal alarms, leading to a reduction in the demands on the audio channel. We hypothesize that a multimodal alarm system can attenuate alarm fatigue and assist the practitioner in recognizing the alarm.

While we have so far not been able to verify this hypothesis, we believe that the experimental protocol we developed to address the research questions here will prove valuable to the multisensory research community and can be applied to future experiments that seek to resolve some of the unanswered questions raised in Section 5. It is also possible that the PoIE, as observed in other experimental contexts, is only manifested in conjunction with speech auditory stimuli, for which the neurophysiological responses are differently affected by the influence of a secondary stimulus modality. This would need to be determined through a separate experiment, employing a speech cue rather than an auditory alarm sound.

Future experiments can employ the PoIE by adding hospital background noise, auditory speech-in-noise tasks, and visual vigilance tasks to test if the hypotheses would still hold while testing clinicians during simulated emergencies requiring clinical pharmacologic intervention. These data will not only inform alarm design and improve patient safety, but have wide-ranging applications to other high consequence industries.

7. ACKNOWLEDGMENT

Colleagues in the Shared Reality Lab, in particular, Francesco Tordini and Jeffrey Blum, provided several rounds of helpful feedback and constructive comments throughout this research, and the technical staff at the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) offered invaluable assistance with our experiments. The authors are most grateful to these individuals. The authors also wish to thank the Vanderbilt University Medical Center, Department of Anesthesiology (especially Drs. Matthew Weinger and Pratik Pandharipande). Vanderbilt University Department of Hearing and Speech Sciences (especially Drs. Wesley Grantham, Ben Hornsby, and Dan Ashmead).

8. REFERENCES

[1] M. H. Otenio, E. Cremer, and E. M. T. Claro, "Noise level in a 222 bed hospital in the 18th health region-pr," *Revista Brasileira de Otorrinolaringologia*, vol. 73, no. 2, pp. 245–250, 2007.

[2] J. H. Kerr, "Warning Devices," *British journal of anaesthesia*, vol. 57, no. 7, pp. 696–708, 1985.

[3] J. Edworthy, "Medical audible alarms: a review," *Journal of the American Medical Informatics Association*, vol. 20, no. 3, pp. 584–589, 2013.

[4] D. Burr, M. S. Banks, and M. C. Morrone, "Auditory dominance over vision in the perception of interval duration," *Experimental Brain Research*, vol. 198, no. 1, p. 49, 2009.

[5] F. E. Block, "For if the trumpet give an uncertain sound, who shall prepare himself to the battle?" *Anesthesia and Analgesia*, vol. 106, no. 2, pp. 357–359, 2008.

[6] J. R. Edworthy, J. J. Schlesinger, R. R. McNeer, M. S. Kristensen, and C. L. Bennett, "Classifying alarms: Seeking durability, credibility, consistency, and simplicity," *Biomedical Instrumentation & Technology*, vol. 51, no. s2, pp. 50–57, 2017.

[7] "Medical electrical equipment, part 18: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems. iec 60601-1-8." *International Electrotechnical Commission*, 2006.

[8] L. B. Talley, J. Hooper, B. Jacobs, C. Guzzetta, R. McCarter, A. Sill, S. Cain, and S. L. Wilson, "Cardiopulmonary monitors and clinically significant events in critically ill children," *Biomedical Instrumentation & Technology*, vol. 45, no. s1, pp. 38–45, 2011.

[9] J. P. Keller, R. Diefes, K. Graham, M. Meyers, and K. Pelczarski, "Why clinical alarms are a top ten hazard: how you can help reduce the risk," *Biomedical Instrumentation & Technology*, vol. 45, no. s1, pp. 17–23, 2011.

[10] M. S. Kristensen, J. Edworthy, and S. Denham, "Alarm fatigue in the perception of medical soundscapes," *European Congress and Exposition on Noise Control Engineering*, pp. 745–750, 2015.

[11] B. Hasanain, A. D. Boyd, J. Edworthy, and M. L. Bolton, "A formal approach to discovering simultaneous additive masking between auditory medical alarms," *Applied Ergonomics*, vol. 58, pp. 500–514, 2017.

[12] P. P. Advisory, "Connecting remote cardiac monitoring issues with care areas." *Pa Patient Saf Authority*, vol. 6, no. 3, pp. 79–83, 2009.

[13] B. Hasanain, A. D. Boyd, and M. L. Bolton, "An approach to model checking the perceptual interactions of medical alarms," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, no. 1. SAGE Publications, 2014, pp. 822–826.

[14] J. J. Schlesinger, "Utilizing multisensory integration to improve psychoacoustic alarm design in the intensive care unit."

[15] M. A. Meredith and B. E. Stein, "Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration," *Journal of neurophysiology*, vol. 56, no. 3, pp. 640–662, 1986.

[16] J. Diamond, P. A. Breslin, N. Doolittle, H. Nagata, and P. Dalton, "Flavor processing: perceptual and cognitive factors in multi-modal integration," *Chemical senses*, vol. 30, no. suppl 1, pp. i232–i233, 2005.

[17] T. Noesselt, S. Tyll, C. N. Boehler, E. Budinger, H.-J. Heinze, and J. Driver, "Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity," *Journal of Neuroscience*, vol. 30, no. 41, pp. 13 609–13 623, 2010.

[18] Y. Visell, B. L. Giordano, G. Millet, and J. R. Cooperstock, "Vibration influences haptic perception of surface compliance during walking," *PLoS one*, vol. 6, no. 3, p. e17697, 2011.

[19] S. A. Gelfand and H. Levitt, *Hearing: An introduction to psychological and physiological acoustics*. Marcel Dekker New York, 1998, vol. 4.

[20] M. Taylor and C. D. Creelman, "Pest: Efficient estimates on probability functions," *The Journal of the Acoustical Society of America*, vol. 41, no. 4A, pp. 782–787, 1967.

[21] M. Taylor, S. Forbes, and C. D. Creelman, "Pest reduces bias in forced choice psychophysics," *The Journal of the Acoustical Society of America*, vol. 74, no. 5, pp. 1367–1374, 1983.

[22] T. N. Cornsweet, "The staircase-method in psychophysics," *The American journal of psychology*, vol. 75, no. 3, pp. 485–491, 1962.

[23] H.-Y. Yao and V. Hayward, "Design and analysis of a recoil-type vibrotactile transducer," *The Journal of the Acoustical Society of America*, vol. 128, no. 2, pp. 619–627, 2010.

PARTICIPATORY DESIGN RESEARCH METHODOLOGIES: A CASE STUDY IN DANCER SONIFICATION

Steven Landry, Myounghoon Jeon

Mind Music Machine Lab
Michigan Technological University
Houghton, Michigan, 49931
{sglandry, mjeon}@mtu.edu

ABSTRACT

Given that embodied interaction is widespread in Human-Computer Interaction, interests on the importance of body movements and emotions are gradually increasing. The present paper describes our process of designing and testing a dancer sonification system using a participatory design research methodology. The end goal of the dancer sonification project is to have dancers generate aesthetically pleasing music in real-time based on their dance gestures, instead of dancing to pre-recorded music. The generated music should reflect both the kinetic activities and affective contents of the dancer's movement. To accomplish these goals, expert dancers and musicians were recruited as domain experts in affective gesture and auditory communication. Much of the dancer sonification literature focuses exclusively on describing the final performance piece or the techniques used to process motion data into auditory control parameters. This paper focuses on the methods we used to identify, select, and test the most appropriate motion to sound mappings for a dancer sonification system.

1. INTRODUCTION

Evidence supports that multimodal interactions increase user engagement with novel interfaces [1]. Therefore, sonification can buttress the connection between the receiver and the information, exploring a new form of art by a synesthetic combination of music and dance. Interactive sonification can be defined as "the use of sound within a tightly closed human-computer interface where the auditory signal provides information about data under analysis, or about the interaction itself, which is useful for refining the activity" [2]. As an interactive sonification technique, parameter mapping [e.g., 3] has often been used, where data features are arbitrarily mapped onto acoustic attributes such as pitch, tempo, timbre, etc.

From this background, we have devised a novel system, immersive Interactive Sonification Platform ("iISoP") for location, movement, and gesture-based interactive sonification research, by leveraging the existing Immersive Visualization Studio (IVS) at Michigan Technological University. The iISoP has been developed for multi-disciplinary research in a verity of fields such as data sonification, gesture interfaces, affective computing, and digital artistic performance. The present paper discusses issues, considerations, and strategies currently implemented in the iISoP's dance-based sonification project, in hopes to spur discussion of applications of more artistic interactions in the sonification community. The selection and fine tuning of motion-to-sound parameter mappings are at the core of any sonification project. Choosing and evaluating these mappings require a network of interdisciplinary team members, each with specific goals and design philosophies that may not always align together. How these decisions are resolved and evaluated are highlighted through a case study in dancer sonification.

1.1. Dancer Sonification

Under normal dance circumstances, the choreographer designs the dance to match specific music. To refer to this type of connection between visual and audio content in multimedia, the term "synchresis" was recently coined [4]. Certain gestures and emotions are utilized to match with specific movements of the musical piece. In the dance-based sonification project of the iISoP, the reverse process is implemented. Music is generated in real-time based on the dance to increase the amount of synchresis between the visual and auditory characteristics of the entire dance experience. The end goal of the dance-based sonification project is to have dancers generate aesthetically pleasing music in real-time based on their dance gestures, instead of dancing to pre-recorded music. The generated music should reflect both the kinetic activities and affective contents of the dancer's movement. The dancer begins to dance, and the sonification system interprets the movements and generates music. The generated music, in turn, influences the way the dancer dances, which is again sonified, leading to a closed loop between the dancer and the system in an interactive manner. To this end, we have collaborated with multidisciplinary teams, involving cognitive scientists, computer scientists, sound designers, dancers, and performing artists.

This dancer sonification project falls in line with previous projects on dance sonification such as the DIEM digital dance system [5], The MEGA project [6], and David Rokeby's Very Nervous System [7]. The iISoP's approach to dance sonfication differs from these past projects in several ways. Our goal for the iISoP is to generate aesthetically pleasing music that is composed of multiple layers or streams of instrumentation. Multiple streams are important to build the body of a musical piece, an important aspect for the immersion of both the dancer and audience. An additional task that previous versions of dance sonifications ignored is affect detection of the dancer, and synthesis of affective content of the gesture sonification that reflects the current state of the dancer.

As with any design research project, the critical aspects to be documented and reported are the methods for which the design is constructed.

https://doi.org/10.21785/icad2017.069

2. SYSTEM CONFIGURATION

The immersive interactive sonification platform (iISoP) is an interactive multimodal system. Figure 1 shows a conceptual diagram of the iISoP system configuration.

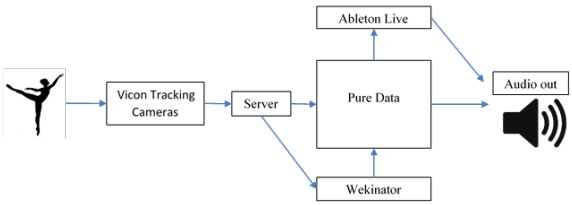


Figure 1: Architecture and data flow of the iISoP system.

The iISoP features a Vicon tracking system utilizing 12 infrared cameras that track specific reflective objects that are strapped to the user's limbs (e.g., wrists and ankles) via Velcro bands. The visual display wall consists of 24 42" monitors controlled by 8 computers that display representations of the tracked objects in real-time. Position, velocity, acceleration, time, proximity to other objects, and holistic affective gestures are recorded and analyzed to generate appropriate sounds (speech, non-speech, music, etc.) based on our own sonification algorithms programmed in Pure Data (real-time graphical dataflow programming environment) [8]. Motion data can also be routed through Wekinator (an open source software tool for real-time interactive machine learning) [9] for machine learning recognition of body postures and gestures. MIDI (Musical Instrument Digital Interface) and OSC (Open Sound Control) messages can also be sent from Pure Data to Ableton Live (music software for MIDI sequencing and music production, creation, and performance) for additional sound synthesis.

3. METHODS

Below are the methodologies in chronological order that we have employed in an attempt at creating an interactive and musically expressive dance-based sonification system.

3.1. Collaboration with performing artist Tony Orrico

As a testbed of our visualization and sonification system, we invited an artist to perform in the iISoP. Tony Orrico is a Chicago based performing artist known for creating large geometric pieces (e.g., Penwald Drawings), using his entire body as an instrument in artistic expression [10]. Orrico demonstrated one of his penwald drawing pieces while wearing sensors that made real-time visualization and sonification. Tony, being a mainly visual artist, had little to contribute to the "sonification" design of the performance, which gave the sound designer full autonomy to choose and implement all parameter mappings. The goal for the sonifications was similar to the goals of the visual presentation: to add a technological aesthetic to the performance piece by reinterpreting and representing the analog movements digitally in real-time.

The sound designer programmed four arbitrary melodies (MIDI format) approximately 1-2 measures long. These MIDI melodies were sent to a "digital bell" sounding MIDI instrument. The instrument and melodies were chosen by the sound designer to convey a particular aesthetic, one that invokes imagery of meditative prayer bells in a monastery mixed with an electronic synthesizer. Melodies were played in an arbitrary order, and the speed of playback was determined by referencing the current velocity of the artist. The purpose of this mapping was to convey to the audience the relationship between the artist's physical movement and the sonic feedback.

For one performance, the artist Tony laid face down on a piece of paper holding graphite pencils in both hands. He pushed off a wall, jetting himself forward on top of the piece. He dragged his graphite pencils along with him; as he writhed his way back to the starting position over and over again, he left behind himself a pictorial history of his motion. While he was drawing these pieces on the paper canvas, his movements created another drawing on the virtual canvas (display wall of monitors) as well as the previously described real-time sonification.

3.1.1 Evaluation

Informal feedback was collected from audience members after the performance was complete. Unfortunately, the audience was generally not impressed with the sonification aspect of the performance. Audience members felt the sonification added little if anything to the overall experience, and it failed as an auditory representation of the performer's physical movements.

On reflection, the sonification "failed" on an implementational level. Melody playback rate was only updated the instant the previous melody finished playing. For instance, if the system happened to update melody speed during a portion of the routine with low velocity, the next melody would be played at a very slow rate (1/4 note = whole notes), which could last for almost 30 seconds. This slow melody would spend 30 seconds describing one instant of the performer's past velocity, growing more irrelevant to the current state of the performer as time passes. If the system happened to update melody speed during the short high-velocity portions of the performer's routine, a quick short melody was triggered, which in the designer's opinion successfully described the movement sonically. Unfortunately, this occurred very rarely. Overall, it was this lack of synchronicity between the activity of the performer and sonic feedback that led to the poor reviews of the sonification. The most obvious movements to the audience (Orrico pushing/jetting from the wall) were often completely ignored by the sonification. This suggested that future versions of the dancer sonification system should include more continuous mapping (rather than triggering discrete melodies) to ensure more synchronicity between motion and sound.

3.2. Dancer interviews

We knew that in order to develop an interactive dance-based sonification system, our background in usability, interactivity, and audio design could only take us so far. With no one on the project having any formal dance training, it was critical to incorporate feedback from domain experts and end users. To this end, we conducted a number of interviews with expert dancers to 1) gather system requirements, 2) evaluate the current versions of our system, and 3) generate novel and intuitive interaction styles and sonification techniques.

Six expert dancers were recruited through local dance performance schools and the local university's Visual and Performing Arts Department. All dancers had at least 10 years of professional dance training. Each semi-structured interview was performed individually, lasting from one to two hours.

The first section of the interview revolved around the expert dancer describing what they would imagine a dancer sonification system to be. This was done before the dancer experienced the current sonification scenario to avoid any anchoring bias. The next section involved the dancer interacting with the system for around 15 minutes while describing their impressions in a "think aloud" fashion. The final section of the interview included a brainstorming session for suggesting modifications and additions to the system, as well as potential applications for the system in other domains.

One interesting theme that came up multiple times through the expert interviews was the importance of valuing the visual aesthetic of the dance over the aesthetic of the sonifications. This has implications over how much control the dancer wishes to have over the sonifications. For instance, dancers would not want to contort their body into odd shapes just to achieve a desired sound. Dancers should also not have to consciously consider every aspect of the sonifications when determining which gesture or posture to perform in sequence. One expert dancer explicitly said "I want 50% of control over the music so I can concentrate on the dance as much as possible". This would require a certain amount of automation on the system side to produce novel and interesting music describing the motion and emotion of the dancer, which accords to the previous experiment [12]. This was in direct conflict with the sound designers associated with the project, who imagined having complete control over every aspect of the sound generation. Musicians place little to no value on the visuals of the gestures, placing all value on the acoustic properties of the sound. From an HCI research perspective, the value is placed on how the user's performance and impression change when the evolutionarily established feedback loop between the dancer and sound is augmented or reversed using technology. In general, each stakeholder has individual goals and philosophies for the project that are at best loosely related, and at worst completely contradictory.

3.3. Visual and Auditory Stimuli Collection

After conducting six interviews, we aggregated general concepts for what expert dancers envisioned how the system should behave. It should first interpret the gestures and affective content of the dance, then create music describing that information. In order to "teach" our system how to perform these tasks, we first had to investigate how humans would accomplish this task. We needed to identify heuristics human composers use to interpret and sonify the motion and emotion of a dance performance. To identify these heuristics, we conducted a small study to collect and analyze visual and auditory stimuli. This study had two goals: 1) to see how and how well non-experts could detect emotion from dance gestures, and 2) to see what type of music or sounds human composers would use to describe dance gestures.

To address the first goal, we invited two expert dancers to submit video recordings of themselves dancing to popular music. These two dancers also participated in the initial interviews. The dancers picked popular songs that represented a particular basic emotion (Anger, Happiness, Sadness, or Content), and performed a dance routine that attempted to portray that particular emotion visually. We then recruited 10 novice participants to watch muted versions of the videos and classify each with the emotion the dancer was attempting to portray. To address the second goal of this exercise, we recruited a music composition class consisting of 10 amateur composers to sonify muted versions of the dancer videos. We gave three specific instructions to the composers. Composers where to: A) re-imagine and recreate the music that the dancers were originally dancing to, B) score the video as if for a film, focusing on capturing the overall mood of the dancer, and C) compose a collection of sounds that describe the kinetic movements of the dancer.

The results of the "guess the emotion" portion of the study suggested that it can be difficult for people to express and interpret emotion through dance gestures alone. There was very little agreement amongst the responses, and self-reported confidence scores were very low. This could be due to a number of factors, but the two most likely explanations of the low accuracy and agreement are 1) communicating emotion through dance is difficult, or 2) non-dancers have difficultly interpreting the intended emotion from dance gestures. Overcoming these obstacles will be critical for embedding automated affect detection algorithms in the iISoP system. The results of the audio stimuli collection portion showed just how infinite the problem space is when considering what type of motion to sound parameter mappings could (or should) be implemented in our dancer sonification system.

Some parameter mapping sonification strategies were consistently used in the majority of audio submissions. Dance gestures that involved rising limbs (raising an arm or leg) were often accompanied with melodies that increased in pitch, and vice versa. Larger body movements were often paired with "larger" sounds (e.g., polyphonic chords, multiple instruments, increase in volume, etc.). Speed of dance gestures was also commonly paired with the speed of the melody (subdivision rate, not BPM of the song). As a note, the project's sound designer was solely responsible for identifying motion-to-sound parameter mappings used in the compositions. This introduces a bias in the type of mappings extracted from the submissions. For instance, mapping height to pitch and velocity to speed was already the intention of the sound designer all along. The same biases certainly unintentionally might filter the information extracted from the expert interviews as well, as the designer could not fully compartmentalize their own goals and philosophy from the interviewee.

3.4. Three new dancer sonification scenarios

We wanted to design a few sonification scenarios leveraging these general strategies used by the human composers from the stimuli validation study. In order to move towards more continuous parameter mapping, we incorporated the real-time graphical programming environment Pure Data into the iISoP architecture. Pure Data afforded us the ability to program virtually any algorithm for real-time parameter mapping sonification. However, designing aesthetically pleasing instruments in Pure Data is time consuming for even the most proficient programmer. In order to leverage the expressivity and control of sound that more conventional DAWS (digital audio work stations) afford to the non-programming population, we included Ableton Live as an alternative means to design and play more aesthetically pleasing instrument sounds.

Two common algorithms we programmed in Pure Data attempted to translate (or map) height to pitch, and velocity to subdivision of generated melodies. For those interested, pictures of the Pure Data subpatches implementing these algorithms are presented below.

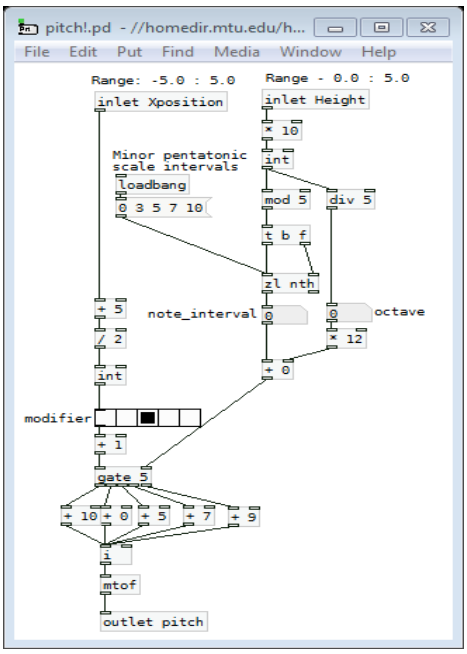


Figure 2: Sonification algorithm for translating position of a tracked object into a MIDI pitch.

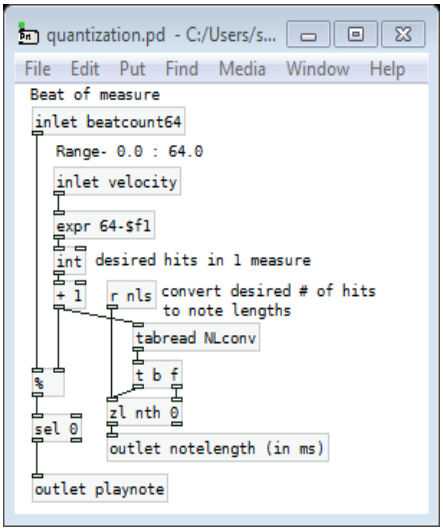


Figure 3: Sonification algorithm for translating velocity of an object to when and how long to play a note within a given measure.

The first of the three newly created scenarios **("A")** focused around a theme of using a user's body as an instrument. Each hand controls independent instruments (melody and percussion). There is a direct mapping between movement speed of that hand and the volume/rate of the arpeggiator for that hand's instrument. Note pitches for the tones are rounded to the nearest note in key, and the onset/duration of notes are quantized in time to the nearest 32nd note subdivision of the tempo. Similar time quantization is used for the percussion instrument using a Euclidean rhythm generator, where the tracked object's current speed determines how many percussion hits are equidistantly distributed across a one measure phrase. The percussion instrument consists of synthetic hi-hat clicks and a bass drum sample. Hand velocity control for the bass drum is scaled down to 1/3 of the rate of the hi-hat clicks to create a syncopated drum rhythm. To provide constant timing cues, a synthetic snare drum is constantly playing on beats two and four of the measure independent of the user's movements. All variable scaling and sound production are done through Pure Data .

The second scenario **("B")** focused around a theme of using a user's body as a DJ's MIDI controller. A very simple 4 measure musical loop was created as a set in the Ableton Live. A number of motion variables were scaled to MIDI range (1-128) using a custom Pure Data patch and routed to through Ableton's MIDI mapping functionality. The user can control a number of parameters controlling the playback of certain instrument tracks or an audio effect applied to the master output. For instance, the right hand's height controls the amount of filter added to a distorted bassline, and the distance between the two hands determines the cutoff frequency of a low pass filter applied to the entire loop playback.

The third scenario **("C")** was a kind of hybrid of the first two themes, where different aspects of the body's overall shape is mapped to a 3 dimensional fader slider controlling the volume balance between 8 pre-made musical loops. Eight musical loops were collected from an online database (all 120 BPM, in the key of C minor, with a length of one, two, or four measures). The musical loops were loaded into a 3D fader object in a custom Pure Data patch for synchronized playback, where each corner of the cube corresponds to one of the eight musical loops. The distance of current position of the fader slider to each of the 8 corners of the cube determines the volume of each of the corresponding musical loops. Six different body shapes (described by distances between the tracked objects) were mapped to the min and max of each of the 3D slider's position variables (X, Y, & Z) using Wekinator. As the user dances or changes poses, the three dimensional fader raises or lowers the volume of each of the 8 musical loops, creating interesting combinations of melodies and rhythms. Note that a sound designer oversaw and configured sonifications of all three scenarios and so, overall sound quality could be similar across the three scenarios.

3.5. Dancer sonification scenario evaluation

In order to evaluate and compare these three newly created scenarios, we conducted a study to evaluate the overall system performance and sonification strategies. Specifically, we wanted to investigate what effect the different interaction styles for each scenario have on user impressions of flow, presence, and immersion in the virtual environment.

Twenty-three novice dancers (M_{age} = 20.3, SD_{age} = 2.1) participated in the study. All participants were recruited from the local university's undergraduate psychology program in exchange for course credit. Eleven participants reported some musical training, and six participants reported some (below 4 years) of formal dance training. Each participant experienced each of the three sonification scenarios for roughly five minutes. This involved the participant exploring and interacting with the system through improvisational dance. Following each scenario, the participant filled out a battery of questionnaires including measures of flow, expressivity, and immesiveness. Participants were also instructed to try and discover and report what motion-to-sound mappings were present in that particular scenario.

Scenario A was reported to have the most "discoverable" or "intuitive" motion-to-sound mappings. Most participants were able to discover at least three of the motion-to-sound mappings regardless of their dance or music demographic backgrounds. Reviews for the overall aesthetics of the sonifications were mixed. Many participants reported the ability to control aspects of the sound that they in reality could not.

Scenario B consistently scored the lowest on the majority of the scales. Many participants reported that the interaction style was confining, not intuitive, and did not encourage the exploration of novel movements. Musicians (especially those who had some experience with digital audio workstations) were more likely to enjoy scenario B and discovered more mappings than non-musicians.

Scenario C was by far the most preferred scenario of the three, and participants suggested it had the most potential for artistic performance applications of the three. Scenario C was also believed to have the most amount of features, even though technically it had the least amount of motion-to-sound mappings. A few participants reported that the interaction style in C was "gratifying". Most participants also mentioned that scenario C's sonifications were the most pleasant sounding of all three scenarios. Participants reported that scenario C's sonifications worked "as a sound representation of the user's movement", the best out of the three scenarios. This was counterintuitive to the designer's expectations, as scenario A was designed to have the most obvious 1:1 mappings between movement activity/location to sound. Scenario C also scored highest with respect to the "the sound helped me understand my movements better" agreement statement.

An interesting finding is that participants often perceived more control of the music than they actually had. For instance, a participant with 4 years of formal dance training reported that he thought he could trigger the synthetic snare drum in scenario A with a sharp deceleration of body movements. In actuality, the snare drum constantly played on beats two and four regardless of user behavior. This was a feature designed to provide familiar temporal cues to the dancer with respect to the tempo and beat of the measure. However, since dancers have been trained to synchronize their movements to these temporal cues, the participant naturally (or unconsciously) synchronized his movements to the automated snare drum. He mistakenly attributed this temporal "coincidence" between motion and sound as a causal relationship. This observation raises additional research questions, such as "what other learned dance behaviors can we leverage to facilitate a richer interaction between user and system?".

Although scenario B was made by a musician for a musician, participants with musical training still preferred the other two scenarios as a whole. Perhaps, a few of the mappings in scenario B were too subtle for non-musicians to notice. In the future, more obvious movements should correspond to more obvious changes in the sonic feedback. Control metaphors used by the designer to control the sound had to be explained to the participants, which suggests these metaphors are not generalizable to others. For instance, the X distance between the hands controlling the low pass filter cutoff frequency was intended to be a metaphor for compressing or stretching the sound as if it was a tangible object.

It was most likely a combination of 1) the clear target goal (isolating an individual loop or achieving a corner position in the 3D fader cube), 2) the challenging method of control through manipulating a body's overall shape, 3) the continuous audio feedback describing the similarity/distance between the current and target body shape, and 4) the obvious and rewarding sound produced once the target shape was achieved that led multiple participants to report that scenario C was "gratifying". Many participants suggested combining aspects of different scenarios for a more expressive performance. Future iterations of the iISoP's dancer sonification phase could combine obvious 1:1 mappings of scenario A and the complex interaction style of scenario C.

In addition to these considerations, more technical aspects of the tracking system need to be revisited. Many of the expert dancers (as well as the non-dancing participants) complained that the objects attached to the ankles and wrists of the user restrict movement, and that more places on the body should be tracked. Before we start adding in more sensors, smaller and more comfortable versions of the sensors need to be designed and tested. The location of hands and feet are only a fraction of the visual information humans use to interpret body posture. Many forms of dance focus on other areas on the body, such as the head, hips, shoulders, elbows, and knees. More data should be collected and used describing the extension angle of joints. There were also struggles with the quality of data from the motion tracking system. Since the dancer's movements often involve spinning, jumping, rolling, the trackable objects worn by the dancer would often be occluded from the vision of the motion tracking cameras, resulting in a large amount of missing data. We also implemented an instantaneous velocity calculation, which resulted in exaggerated jumps in the reported velocity/acceleration data. We will switch to using a rolling average instead to smooth out the data in future scenarios.

3.6. Dancer Workshops (ongoing)

Another set of scenarios are currently being developed based on the feedback described in the dancer sonification evaluation study. These scenarios will be designed and evaluated during multiple workshops in collaboration with invited expert dancers. Dancers will present during the programming of Pure Data patches to help inform the programmer of appropriate scaling values when translating motion to sound parameters. It is expected that once the dancers have a general feeling of the types of algorithms implemented in Pure Data (through this interactive process), they will be able to suggest more creative potential parameter mappings than in previous brainstorming sessions.

The main direction of the new scenarios is to give the user the ability to control the overall structure and flow of a song, opposed to the static set of instruments featured in the first three scenarios. Since the majority of popular music is structured into repeated sections (intro, verse, chorus, bridge, outro, etc.), giving the user the ability to switch between these sections is another step to accomplishing the end goal of the dancer sonification system. Programmatically, this suggests that sets of premade instruments should be available at all times to the user. The user should also be able to activate, mute, or modify the pre-made instrument sets through specific gestures, locations in the room, or through intervening actions taken by the iISoP system based on a rolling average of the "quality of movement" of the dancer. The software "Eyesweb" [13] shows promise in calculating and routing automated "quality of movement" analysis to our sonification software. The quality of motion is based on Laban Movement Analysis, and affords us a much better description of dance gestures than simple velocity and distance calculations.

4. DISCUSSION AND DESIGN CONSIDERATIONS

From all of these works, we have learned how valuable domain expert feedback can be for designing a dancer sonification system. We have also learned how difficult it can be to integrate competing ideas from different stakeholders. We are also starting to unpack exactly how the interaction style and sonification methods can influence users' feelings of presence and immersion in virtual environments. Simply affording the users the ability to control certain aspects of the auditory display does not guarantee interactivity, nor does it guarantee that the users feel "immersed" in the virtual environment. More features and more complex mappings do not equate to "richer" interactivity. Users do not have to completely understand every motion-to-sound mappings in order to express themselves artistically. There are certain aspects of the auditory display that users expect to be able to control, and are disappointed when the system does not conform to their expectations. However, what is perhaps more useful is knowing which aspects of the auditory display can be automated to ensure the music is aesthetically pleasing while still depending on the user's input. These automated strategies alleviate the users' workload to focus on the more creative aspects and dance and composition instead of "trivial" aspects such as specific MIDI pitches and note lengths. We have also learned that the efficacy of different control metaphors is heavily dependent on the user's personal experiences.

Creating a balance between user control and system automation is difficult. Enough automation is necessary to ensure the sonic output of the system is pleasant and structured, like typical popular music. However, embedding too much automation begins to deteriorate the perceived connection between the gestures and music. Giving the user too much control of the sonic output has negative effects on the cognitive flow of the dancer, and the physical flow of the dance performance. A certain amount of stochasticity in the mappings or sonic output may be necessary to keep the music from becoming repetitive. It is important to include what we know about how expert humans compose music (heuristics) in the design of sonification algorithms. Keeping notes in key and using a constant BPM are obvious composition heuristics, as is spreading out audio streams over wide frequency spectrum (e.g., bass, melody, lead, percussion). Designers must keep in mind that the music must still sound musical, and the dance must still resemble dance, otherwise it is no longer a dancer sonification system.

5. REFERENCES

[1] Hermann, T., & Hunt, A. (2004). The discipline of interactive sonification. In Proceedings of the International Workshop on Interactive Sonification.

[2] Hermann, T. (2008). Taxonomy and definitions for sonification and auditory display. In Proceedings of the 14th International Conference on Auditory Display (ICAD 2008).

[3] Dubus, G., & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. PloS one, 8(12), e82491.

[4] Bencina, R., Wilde, D., & Langley, S. (2008, June). Gesture≈ Sound Experiments: Process and Mappings. In NIME (pp. 197-202).

[5] Siegel, W., & Jacobsen, J. (1998). The challenges of interactive dance: An overview and case study. Computer Music Journal, 22(4), 29-43.

[6] Camurri, A., De Poli, G., Friberg, A., Leman, M., & Volpe, G. (2005). The MEGA project: Analysis and synthesis of multisensory expressive gesture in performing art applications. Journal of New Music Research, 34(1), 5-21.

[7] Rokeby, D. (1998). The construction of experience: Interface as content. Digital Illusion: Entertaining the future with high technology, 27-48.

[8] Puckette, M. (1996). Pure Data: another integrated computer music environment. Proceedings of the second intercollege computer music concerts, 37-41.

[9] Fiebrink, R., & Cook, P. R. (2010). The Wekinator: a system for real-time, interactive machine learning in music. In Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)(Utrecht).

[10] Jeon, M., Landry, S., Ryan, J. D., & Walker, J. W. (2014, November). Technologies expand aesthetic dimensions: Visualization and sonification of embodied Penwald drawings. In International Conference on Arts and Technology (pp. 69-76). Springer International Publishing.

[11] Walker, J., Smith, M. T., & Jeon, M. (2015, August). Interactive Sonification Markup Language (ISML) for Efficient Motion-Sound Mappings. In International Conference on Human-Computer Interaction (pp. 385-394). Springer International Publishing.

[12] Jeon, M., Winton, R. J., Henry, A. G., Oh, S., Bruce, C. M., & Walker, B. N. (2013, July). Designing interactive sonification for live aquarium exhibits. In International Conference on Human-Computer Interaction (pp. 332-336). Springer Berlin Heidelberg.

[13] Camurri, A., Hashimoto, S., Ricchetti, M., Ricci, A., Suzuki, K., Trocca, R., & Volpe, G. (2000). Eyesweb: Toward gesture and affect recognition in interactive dance and music systems. Computer Music Journal, 24(1), 57-69.

Paper Session 7
Computing

COMPUTATIONAL DESIGNING OF AUDITORY ENVIRONMENTS

David Worrall

Audio Arts and Acoustics Department,
Columbia College Chicago,
600 South Michigan Avenue Chicago, IL 60605, USA.
dworrall@colum.edu

ABSTRACT

This paper is a call for sonification designers to adapt their representational practices from that of designing objects for auditory engagement to the construction of systems of formally described relationships that define the 'state space' from which streams of such objects can be drawn. This shift from the crafting individual sonic objects and streams to defining dynamical space of design possibilities we call 'computational designing'. Such sonification model spaces are inaudible, heard only through its instances, or the manifestations of particular *trajectories* through the space.

Approaching the design of auditory displays as computational tasks poses both considerable challenges and opportunities. These challenges are often understood to be technical, requiring scripting or programming skills, however the main challenge lies in computational design thinking which is not best understood as the extension of established designing processes.

The intellectual foundations of computational designing rest at the confluence of multiple fields ranging from mathematics, computer science and systems science to biology, psychophysical and cognitive perception, social science, music theory and philosophy. This paper outlines the fundamental concepts of computational design thinking based on seminal ideas from these fields and explores how they it might be applied to the construction of models for synthesized auditory environments.

1. INTRODUCTION

How sonification designers attempt to achieve an effective communication solution with a sonification design task is affected by many things, including the imagined range of possible solutions for the design (the state space), which in turn, is affected by the tools and methodologies employed, and the skills applied in using them. Attempting to evaluate the extent to which the solution is optimal, and how effective it is in achieving a stated goal, remain open questions because the connection between the individual decisions that are made in the designing process and the way they interact in the solution space are non-linear, at best open to interpretation and at worse a collection of individualized black-box heuristics. Such a description is rarely controversial, even by those calling for robust evaluation and scientific comparison of sonification methods, as it is understood that:

> *In the context of data exploration, what can be potentially learnt from a sonification is unknown, or at least not defined properly, and therefore it is very difficult to specify an objective performance measure.* [1]

Design is a messy business and the relationship between decisions made in the process of 'tweaking' the contribution of individual parameters in the final result is rarely the sum of simple linear combinations. So, being able to evaluate the effectiveness of a sonification for a clearly defined purpose is not the same as being able to determine what aspects of the sonification are responsible for or contribute to that effectiveness. This is no different in form to being able to construct a neural-network to simulate an observable behavior without being able to understand the complexity of the individual contributions to the network itself.

2. CURRENT DATA SONIFICATION MODELS

Various methods have been developed to use synthesized sounds for the purpose of sonifying data.[1] Perhaps the simplest, or at least the most direct is *audification*, a "direct translation of a data waveform to the audible domain." [2]. Audification may be applicable as a sonification technique to many data sets that have an equally spaced metric in at least one dimension. It is most easily applied to those that exhibit oscillatory time series characteristics, although this is not a requirement.

In Model-Based Sonification [3][4], a variable of the dataset to be sonified is assigned to some structural properties of a component (elasticity, hardness etc) of the model. A user interacting with this model via 'messages'– virtual beaters, scrapers, etc.–causes it to 'resonate'. The resulting sound is thus determined by the way the data integrates through the model under excitation by the messages. By virtually beating, plucking, blowing and scraping the model (the characteristics of the dataset) are available to the listener in a way analogous to the way that the material and structural characteristics of a physical object are available to listeners who beat, pluck, blow or scrape it.

Parametric-Mapping Sonification (PMson) is the most widely used sonification technique for representing high-dimensional data as sound. Typically, data dimensions are mapped to sound parameters: either to physical (frequency, amplitude), psychophysical (pitch, loudness) or perceptually coherent complexes (timbre, rhythm). At its simplest, PMson uses homomorphic mapping in which the changes in one

https://doi.org/10.21785/icad2017.059

[1] Data sonification is the acoustic representation of data for relational interpretation by listeners, for the purpose of increasing their knowledge of the source from which the data was acquired.

dimension of the auditory space tracks changes in a variable in the dataset, with as few mediating translations as are necessary for comprehension [5]. There is a kind of degenerate approach to PMson that uses sampled rather than synthesized sound objects. Such an approach limits the parameter-mapping to gross sound parameter control without the options of sophisticated modification that synthesizing sound objects provides. There are some instances where a combination of the two techniques is appropriate.

PMson is the closest data sonification method to the traditional musical model of an abstract score (the data) rendered into sound. Largely because of the number of data dimensions it can potentially carry, PMson remains the most flexible and potent method for embedding higher-order hierarchical percepts such as timbre and meter, which is necessary to overcome some of its limitations, such as the parameter-mapping problem which arises from the non-orthogonal nature of the hearing system [6][7][8][9]. In order for this embedding to occur, we need to develop tools and techniques to compute user navigable auditory environments (sonifications) from a conceptual infrastructure in which are embedded cognitive and psychoacoustic supports for the basic techniques.

1. SONIFICATION DESIGNING AND MUSIC COMPOSITION

Many sonification designers are also music composers who have spent hundreds of hours, not unusually from a very early age, learning to listen intently and experiment intelligently. They have learnt to compose sound combinations that either crystalize into harmonically coherent complexes (harmonies) or individual streams of sound (musical lines) that maintain various degrees of independence from each other when sounded simultaneously (counterpoint). Many composers also undertake studies in orchestration, which involves both learning how, over hundreds of years, composers of different of styles of music have scored harmonic and contrapuntal ideas to different effect, as well learning how to score abstract musical structures for performance by a wide variety musical instruments for their own compositions.

Composers thus bring to the task of data sonification a plethora of skills and experience that have been black-boxed into their individually developed Heuristic Auditory Environment Synthesis (HAES)[2] techniques. Many of the techniques they have acquired have their foundations in psychoacoustic principles even though they were developed, often over centuries, by composers who listen, rather than by psycho-acousticians or cognitive psychologists. There are a large number of such "rules-of-thumb': those for maintaining linear integrity by avoiding parallel fifths and octaves, for example, or for where in the tessitura of various instruments of different timbres the attack transient onset-times are similar, thus affording the smooth dove-tailing of multiple-octave scalar sweeps, or how to adjust relative onset times to ensure simultaneities (chords) are heard as such.

It has become increasingly recognized that design is not free expression and is not scientific engineering; it is in-between (a 'third culture'), sometimes fluxing more towards one than the other; the application of scientific *and other organized knowledge* to practical tasks. Today, these increasingly artificial or designed worlds are also digital. In fact, they rely on digitization. Yet, just as the designed world is rapidly replacing the natural world as the principal mode of (humans) being in the world, it is widely recognized that many design problems are ill-defined.

> *They are not the same as the "puzzles" that scientists, mathematicians, and other scholars set themselves. They are not problems for which all the necessary information is, or ever can be, available to the problem-solver. They are therefore not susceptible to exhaustive analysis, and there can never be a guarantee that "correct" solution-focused strategy is clearly preferable to go on analyzing "the problem," but the designer's task is to produce "the solution."* [10]

There are some efforts to make sonification a well-defined scientific method, i.e. the "data-dependent generation of sound that uses systematic, objective and reproducible transforms" [11]. However, the act of designerly producing a sonification obscures its methods of production, leaving an object that is resistant to analysis, or at best not conducive to it. So, this 'covering up' of the design process does not assist the development of better sonification practices. Nor does it promote the reproducibility requirement which has been declared by some as necessary for sonifications to be called scientific:

> *"… given the same data and identical interactions (or triggers) the resulting sound has to be structurally identical."*
>
> *The definition claims reproducibility. This may not strictly be achieved for several reasons: … The use of the term "structurally identical" in the definition aims to weaken the stronger claim of sample-based identity. Sample-based identity is not necessary, yet all possible psychophysical tests should come to identical conclusions.* [1]

The above discussion suggests a need for sonification design decisions to be derived from more explicit criteria than the educated responses of designers with black-box skills. It is not just a question of the importance of user-centered interactive design replacing fixed-format audio files. What is needed is a deeper and more encompassing codification of the design process itself, including all the relevant black-box contents that go into heuristic decision-making processes, the translation of historical and current knowledge into computer-actionable knowledge banks and the ability for such systems to both suggest novel solutions and to learn and adapt to individual designers' styles. This is a non-trivial task but one that is being attempted in allied disciplines such as audio game engine design.

3. COMPUTATIONAL DESIGNING[3]

The dominant mode of utilizing computers for audio design and production today is computerization or compilation: sonic objects or processes that have been conceptualized in

[2] A term I invent here to distinguish the process from Computational Auditory Environment Synthesis (CAES), discussed later. The term Auditory Environment Synthesis is preferred to Auditory Scene Synthesis (ASS) as it emphasizes the three-dimensional surrounding nature of sonically being in the world.

[3] This section relies heavily on ideas developed through the synthesis of multiple authors in [12].

the designer's mind are recorded, manipulated and/or stored in a computer using a Digital Audio Workstation (DAW). Typically, a DAW consists of a computer with sufficient storage and processing speed to be able to mix multiple simultaneous channels of audio to which pre-programmed processes are applied (often in the form of 'plugins') by software (such as *Protools*, *Abelton Live, Logic* and *Reaper),* which are used to edit, mix, transform, and store or simultaneously play back those audio channels for direct audition.

From a design perspective, this computer-aided, compilation approach does not allow the designer to take advantage of the *computational* power of the computer in the design process itself.

> *The manifest form - that which appears - is the result of a computational interaction between internal rules and external (morphogenetic) pressures that, themselves, originate in other adjacent forms (ecology). The (pre-concrete) internal rules comprise, in their activity, an embedded form, what is today clearly understood and described by the term algorithm.* [12]

Expressed simply, computational designing employs computation in the design process itself to *deduce* and place elements, for example, real-time synthesized sounds including microsound responsiveness to situational criteria, user input and the like. Computational design is best understood when compared to computerized (or computer-aided) design, in which the computer is used to *compile* and arrange fixed design elements such as transforming pre-recorded sound samples to better fit the specific situation in which they are to be used.

Computerized design is based on a data model, whereas computational design relies on a procedural model. Computational design involves the processing of information and interactions between elements that constitute a specific environment. The final appearance of these elements, whether they be game objects or sonified information derived from data, is rendered from the processing of intrinsic properties, such as the specific values of data points, important information beacons, or extrinsic properties such as the positional rendering of the object in the acoustic environment in which it is being placed, taking into account the effect (salience, occlusion[4] etc) of other objects that have already been or will be placed there. Computation provides a conceptual framework for highlighting the data being rendered according to the importance placed on it at the time by the designer and the interacting user. As a design methodology, whereas computer-aided design begins with *objects* (such as sound samples) and adapt them to specific situations, computational designs start with *elemental properties* of objects (as synthesis parameters) and environmental influences, and use generative rules to formulate (or proceduralize[5]) the specific objects in the specific environment into which they are placed.

[4] The salience of a sound is its attention-grabbing or distinctiveness[17]; occlusion refers to the virtual hiding or masking of a sound by others. These characteristics can be altered using signal processing techniques such as filtering and reverberation.

[5] The terms Procedural Design and Computational Design are frequently used interchangeably. Procedural Design is the more general term. When they are computed, Procedural Designs become Computational.

Most of the work of the computational designer involves explicitly defining and editing the definition of sets of variables such as psychoacoustic parameters and constraints. In generating specific solutions, logical operations on these sets and their (often dynamically generated) subsets are performed without the designer necessarily being able to conceptualize the full formal implications of their relationships. This can be a positive consequence of such abstraction as it can produce state-space solutions that might not have been intuited, considered or imagined using non-computational approaches. Because the designer is freed from the requirement to produce a single 'masterful' solution, many instantiations can be produced and then evaluated for their effectiveness, in-keeping with the goals of the SonEx project [1].

3.1. Most music is composed procedurally

Adapting Schön and Wiggins' three types of seeing for architectural designing [13] to listening, we have

1. Literal aural apprehension of auditory objects–sensorially-led appreciation.
2. Appreciative judgements of quality (tone, texture, timbre, pitch and, duration), discovered by more intellectually–led reduced listening, and 'deep' pattern recognition.
3. Apprehension of auditory state-space gestalts which are not instantaneously heard or imagined. Such higher-order designs are well known to be important in music and include things like beat, pulse, meter, swing, scales, modes, chords, etc;–higher level perceptual groupings that afford memory retention, comparison and contrast.

Most composing undertakes these steps in reverse: General conceptual ideas and gestalts like melodies and harmonies precede thoughts about instrumentation and scoring. We can observe this in music that can be performed on multiple instruments: it is the structural organization of the sounds which becomes primary, not the sensation of the tones used to "carry" the organization. This principle is also exemplified by the fact that, historically, many large-scale instrumental compositions were composed first into a keyboard score, with specific instrumental orchestration of these ideas being undertaken as separate and later processes.

Generative music composition has a long tradition, pre-dating the invention of the digital computer by several hundred years. With the advent of the computer, it quickly became formalized [14]. Equipping machines with the ability to play a vitally important role in achieving musical tasks, such as composition, improvisation, and accompaniment, is now an important development in innovative musical practices.

3.2. Digital sound synthesis

In addition to their use in compiling audio, computers can also be used to digitally generate or synthesize new sounds. Such computation has had a profound impact on our ability to produce conceptually simple sounds such as sine tones as well as sounds of great complexity such as stochastically controlled microsound-grains as well as complex sounds using simulations of physical resonators such as acoustic musical instruments. Early work by pioneering composers and engineers such as Xenakis, Risset, Chowning, Mathews, and Moorer, was seminal in establishing both tools and techniques for exploring and understanding the dynamic

nature of natural sounds and thus the design of 'lively' and interesting computer-generated tones and the multi-reverberant environments in which they are formed.

The synthesis of individual sounds is an intensely computational process and algorithms for doing so using granular, physical models, fixed-waveform and other hybrid techniques are available today in a number of mature, well-honed applications such as *Csound*, *Supercollider* and *Pure Data*. Though not sufficient in themselves to design sophisticated auditory environments, such programs already contain the necessary sound-synthesis tools for supporting a computational approach.

3.3. Some advantages of using a computational approach for designing auditory environments.

While there may be some circumstances in which auditory icons[6] and earcons[7] might be sufficient for creating a useful sonic environment, having to rely on them, with some general sonic smudging to attempt to smooth over the cutting-and-pasting, is not a recipe for the development of more sophisticated and responsive sonifications which are clearly needed as the understanding of the psychoacoustic and cognitive correlates of data sonification increases. As the complexity of an environment increases and/or the number of object in it increases, computational load of rendering all auditory objects becomes a critical defining feature of how many such objects can be rendered. In a sample-based model, either the library of different samples for an increasing number of sonic objects has to be the generated and compiled, or the processing requirements for modifying a smaller subset of each object has to be increased.

In a computational design model, the rendered form of both individual auditory objects and general environmental factors–and the ways in which they interact–can be dynamic and highly flexible. This increases variety and reduces the reliance on the modification of decisions that need to be made in advance when using a samples model. For example, whereas sound samples might need to be modified according to situational salience requirements, in a computational model, the salience of objects can be made a feature of the synthesis of the objects themselves. In a sonic environment which is responsive to user-directed interests, for instance, this affords the production, of a better balance between local and global reverberation requirements, resulting in a deeper sense of environmental continuity and sound-object integration.

Another advantage of using a procedural approach is that, as the level of detail needed and the number of objects increases in the rendered auditory environment, the overall computation time involved relative to that required by manipulating samples, significantly reduces.

4. A CASE STUDY: AUDIO IN COMPUTER GAMES

A game world is composed of discrete entities (objects) that have a set of properties (name, visual appearance, location, sound signature etc) and a set of behaviors and sounds associated with those behaviors as they interact in the world, and/or the world in which they exist interacts with them. Early games used generative techniques to produce sound-effects and musical accompaniment before computers became powerful enough to use sampling technology to deliver 'realistic' recorded sounds. Because a sound recording captures the digital signal of a single instance of a sound and not its dynamic behavior, many clever tricks have been developed to blend, layer, filter, interpolate and time-warp these singular sound 'instances'. What this has meant for game audio over the last two decades is that overwhelmingly, the auditory design approach has remained *event based* (i.e using sound samples) despite Farnell's prediction that "procedural sound is set to make a huge comeback":

> *Traditional game audio binds each action to an event which triggers a sound sample. Some real-time modifications can be applied such as distance damping, or combining alternative samples in a random or granular way to get more variation. But none of these techniques is able to map the underlying physical behavior an object and its sound. …an equivalent event-based graphical game would only be a series of static photographs. …*
> *In software engineering terms, games audio is badly coupled and inconclusive …* [18].

This event-based approach causes timing problems such as the aligning sample loops with game movements, thus limiting visual and/or audio actions to a predetermined duration and raises the need to devise new methods, such as randomly selecting samples and sub-samples, to alleviate the repetitive quality of a limited data set. What is somewhat surprising is that although the use of computational generation of dynamic sound objects is being introduced to game design, the approach has not (yet?) made the huge comeback Farnell predicted. The most common usages seeem to be in continuous and enviromental sounds:

> *Our procedural audio synthesis technologies use minimal amount of samples to generate a great variety of high-quality sound events. More importantly, our synthesis engine comes with specially-designed controls for intuitive and instant sound manipulation. Currently available technologies include wind, rain, fire, electric sparks and space ambience synthesizers. Some more are coming*. [19]

Joe Cancellaro, composer and game design educator informally described the current state in this way:

> (procedural audio) *did not make a huge comeback. Sampled audio is pervasive in the game industry. What has changed significantly are the audio engines that deliver the samples, processing them (samples) in real-time, and in many cases altering the original sample. This could be seen as procedural but is it not really. There are examples where procedural sound is used in games, but it turns out it is less effective, emotionally, than samples. It lacks a "human" touch. For big pads and sustained environments procedural could work, but even then, I would go with a composed sample and let the non-linear audio engine chew it up and make something contained but random. It's all Stravinsky at that point. Shackles!* [20]

[6] Auditory icons are "sounds which mimic everyday non-speech sounds that we might be familiar with from our everyday experience of the real world" [15]

[7] Earcons are "…short, structured musical phrases that can be parameterized to communicate information in an Auditory Display." [16]

Overtly emotive sounds have a more function functional role in developing and enhancing satisfying game experiences than they do in data sonification tasks. So, it will be interesting to observe if, as game worlds increase in size and complexity, the time critical interactivity requirements that gamers demand will eventually force designers to rely more on computational sound design. Perhaps the increasing computer power of the platforms on which games are delivered will be sufficient to permit game sound designers to continue using sound samples as a means of delivering satisfying user experiences in the foreseeable future.

5. NEW DIRECTIONS IN COMPUTATIONAL DESIGNING OF AUDITORY ENVIRONMENTS

This paper has its origins in thinking about The Mapping-Problem and a developing sense that, although there is a general understanding of the Problem there seemed be no solutions to circumventing it that don't involve the compilation and integration of a large body of psychoacoustic and cognitive knowledge in a sonification design framework.

Incorporating task-based analysis into design criteria is increasingly understood to be an important step in developing effective data sonifications. At the same time, there is a need to develop larger solution state-spaces with an increased number of dimensions, incorporating both microsound and gestural levels and the intelligence to form multiple mapping solutions from correspondences between them which produce both highly dynamic and situationally responsive individual sound objects and higher-order extra-objective perceptual experiences such as swing, which is produced by systematically modulating the temporal flow.

The ability for designers, and ultimately uses, to adapt a data sonification to their aural developing skills is also important. This is supported by Jean Piaget's observation of an interesting aspect of the relationship between representational and perceptual space. During the development of an understanding of a representational space through experimentation (or 'play'), representational activity is 'reflected' or 'projected back' on to perceptual activity [21] as exemplified in the way understanding of musical structures affects the way one perceives musical affect. This observation supports the hypothesis that for a listener, there is dynamic relationship between their critical listening skills, and the sonic complexity and variability of a sound mapping that can be understood. This emphasizes both the need to have auditory designs which are responsive and/or able to accommodate different and developing listening skills, but also the need to develop the integration of the development of listening and sound mapping experiences in early childhood education curricula.

If perception is not solely a bottom-up process but incorporates inherited or acquired response biases related to stimulus characteristics and sensory systems that have evolved for multiple, integrated representations of certain features which are meaningful to the perceiver rather than for just single one-to-one reproductions of the physical world, it makes sense to generate multiple sonifications of any particular data set according to the user's developing sense of meaningfulness. In order to accomplish this, I suggest it is necessary for sound designers interested in creating dynamic auditory environments to shift their design representation and thinking efforts from creating bespoke hand-crafted solutions using just their own black-boxes, to a dynamic system model approach in which design activity is supported by extended state spaces that incorporate psychoacoustic and cognitively-informed transforms.

Because of the breadth and depth of detail required, sets of such transforms need to be developed collaboratively using agreed-upon standards. A community-based approach to developing these resources should go some way towards increasing the computational power of the computer as a design tool to *deduce* a wider variety of more effective auditory designs.

6. REFERENCES

[1] Degara, N., F.Nagel, and T.Hermann, 2013. SonEX: An evaluation exchange framework for reproducible sonification, in *Proceedings of the International Conference on Auditory Display (ICAD)*, pp. 167-174.

[2] Kramer, G., 1994. Some organizing principles for representing data with sound. In G. Kramer (ed.). 1994. *Auditory display: Sonification, Audification, and Auditory Interfaces*. Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, Volume XVIII. Reading, MA: Addison Wesley Publishing Company, p186.

[3] Hermann, T., & H. Ritter, 1999. Listen to your data: Model-Based sonification for data analysis. *Lasker, G. E. (Ed.), Advances in intelligent computing and multimedia systems*, 189-194.

[4] Hermann, T., 2002, *Sonification for exploratory data analysis*, Ph.D. thesis, Bielefeld University, Bielefeld, Germany.

[5] Kramer, G., 1994. An introduction to auditory display. In G. Kramer (ed.). 1994. *Auditory display: Sonification, audification, and auditory interfaces*. Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, Volume XVIII. Reading, MA: Addison Wesley Publishing Company, p26.

[6] Worrall, D., 2014. Can micro-gestural inflections be used to improve the soniculatory effectiveness of parameter papping sonifications? *Organised Sound, Volume 19, Special Issue 01*, Cambridge University Press, pp 52-59.

[7] Worrall, D., 2013. Understanding the need for micro-gestural inflections in parameter-papping sonification. *Proceedings of the 19th International Conference on Auditory Display*, Łódź, Poland July 6-10, pp 197-204.

[8] Worrall, D., 2011. A method for developing an improved mapping model for data sonification. *Proceedings of the 17th International Conference on Auditory Display*, Budapest, Hungary, June 20-24.

[9] Worrall, D., 2010. Parameter mapping sonic articulation and the perceiving body. *Proceedings of the 16th International Conference on Auditory Display*, Washington, D.C, USA, June 9-15.

[10] Cross, N., 1982. Designerly ways of knowing. *Design Issues*, 3(4), 221-7.

[11] Hermann, T., 2008. Taxonomy and definitions for sonification and auditory display, in *Proceedings of the 14th International Conference on Auditory Display (ICAD 2008)*, P. Susini and O. Warusfel, Eds. IRCAM, Paris, France.

[12] Kwinter, S., 2008. Far from equilibrium: Essays on technology and design culture, Actar, Barcelona, p 10. as quoted in A. Menges and S. Ahlquist, *Computational design thinking*. Wiley and Sons, 2011, p 147.

[13] Schön, D. and G. Wiggins, 1992. Kinds of seeing and their functions in designing, in *Design Studies*. pp 135-56.

[14] Xenakis, I., 1971. *Formalized music: Thought and mathematics in music*. Indiana: Indiana University Press.

[15] Brazil, E. and M. Fernström, 2011. Auditory icons, in T. Hermann, T., A. Hunt and J. Neuhoff (Eds.) (Eds.), The *Sonification Handbook*. Berlin: Logos Verlag, Ch 13, p325-338.

[16] McGookin, D and S Brewster, 2011. Earcons, in T. Hermann, T., A. Hunt and J. Neuhoff (Eds.), *The Sonification Handbook*, Berlin, Germany: Logos Publishing House, Ch14, pp. 339-361.

[17] Delmotte, V., 2012. *Computational auditory saliency*, Ph.D. Thesis, Georgia Institute of Technology, Atlanta.

[18] Farnell, A. 2010. Sound Design. The MIT Press., p 318.

[19] https://lesound.io/technologies/

[20] Cancellaro, J., 2017. (Chair, Interactive Arts and Media Department, School of Media Arts, Columbia College Chicago.) Personal communication.

[21] Piaget, J and B. Inhelder, 1956. trans FJ Langdon and JL Lunzer, *The Child's Conception of Space*, Routledge & Kegan Paul (London; New York), p 4.

PARALLEL COMPUTING OF PARTICLE TRAJECTORY SONIFICATION TO ENABLE REAL-TIME INTERACTIVITY

Jiajun Yang

Ambient Intelligence Group
CITEC, Bielefeld University
Bielefeld, Germany
jyang@techfak.uni-bielefeld.de

Thomas Hermann

Ambient Intelligence Group
CITEC, Bielefeld University
Bielefeld, Germany
thermann@techfak.uni-bielefeld.de

ABSTRACT

In this paper, we revisit, explore and extend the Particle Trajectory Sonification (PTS) model, which supports cluster analysis of high-dimensional data by probing a model space with virtual particles which are 'gravitationally' attracted to a mode of the dataset's potential function. The particles' kinetic energy progression of as function of time adds directly to a signal which constitutes the sonification. The exponential increase in computation power since its conception in 1999 enables now for the first time to investigate real-time interactivity in such complex interweaved dynamic sonification models. We speeded up the computation of the PTS model with (i) data optimization via vector quantization, and (ii) parallel computing via OpenCL. We investigated the performance of sonifying high-dimensional complex data under different approaches. The results show a substantial increase in speed when applying vector quantization and parallelism with CPU. GPU parallelism provided a substantial speedup for very large number of particles comparing to using CPU but did not show enough benefit for a low number of particles due to copying overhead. A hybrid OpenCL implementation is presented to maximize the benefits of both worlds.

1. INTRODUCTION

Model-Based Sonification (MBS) is a sonification technique that involves (usually high-dimensional) data into the definition of dynamic systems which behave according to given laws of motion. Interaction modes (e.g. shaking, knocking) provide the means for interactively exploring features of the model (and thus coherences within the underlying data) as they result in a dynamic system behavior that directly contributes to a sound signal, i.e. the sonification [1].

Let's take an example of studying how a new percussive instrument works and sounds like. We can visually observe the construction of the object, then touch over the instrument and feel the materials of different parts. Then we can give it excitations through various physical actions, e.g. tapping, hitting, scratching, to test how each part of the instrument sounds like. To understand the sound it can provide, these excitation will also accompany with different dynamics, e.g. tapping it gently or hitting it hard. Furthermore, we can pick up other objects such as a drumstick to interact with the instrument to find out more acoustic possibilities. Through these, one can gain a good understanding of the functionality of the instrument.

Model-Based Sonification applies the similar analogy to the sonification for data mining. For instance, a dataset has its own intrinsic features which are unknown. However, we can define a dynamic model to bridge between the abstract data and the domain of acoustical systems, which are nothing but physical systems which react dynamically. Assuming that a model designer also implemented modes of interactions, the user can then also provide excitations to the data-driven configuration in model space and in turn the sonification model generates the acoustic response as immediate feedback to the user, whatever the unforeseeable interaction will be. The data imprints acoustic model properties which holistically and characteristically manifest in sound properties.

A Model-Based Sonification design usually consists of the following steps:

1. The *setup phase* defines how the data configure elements of the model (which typically exist in a model space). The data themselves are neutral and abstract, merely numbers. The setup grants them physical properties, to establish a *dynamic* system with internal degrees of freedom. For example, each data point can be considered a point mass in a d-dimensional space. Or each data point might be considered as a mass-spring system with inherent attributes such as mass, stiffness, thus an element that can oscillate and exhibit acoustic behaviors.
2. The *model dynamic* sets the equations of motion that ultimately determines the temporal evolution of the system and thus the sound. It can make use of physical principles such as inertia, friction, propagation, and energy conservation.
3. The *excitation phase* is the core part of the interaction to trigger auditory display. In this phase, specific excitation methods need to be defined, whether it is through simple Mouse clicks or complex tangible/physical interactions. This phase is also closely tied with the *model dynamic*.
4. The *link variables* are model-specific features that can be calculated at any time and whose temporal progression usually delivers directly the sound signal.

In data mining, *Exploratory Data Analysis* defines the process to acquire understanding of data and detect patterns when an explicit knowledge of the data is absent. In the 1970s, Tukey established many famous visualizations of uni- and multivariate data,

https://doi.org/10.21785/icad2017.022

e.g. leaf plots, histogram. While they are still the predominant methods in data visualization, visualizing high-dimensional data is not always an easy job. For example, parallel coordinate [2] plots are commonly used to visualising multivariate data by juxtapositioning each data dimension along one axis. However, important patterns and features can be difficult to distinguish when the data space is dense [3]. Scatterplot matrices are particular useful when exploring the correlation between any two variables in the data space, but they lack the ability to examine the holistic feature of the data. The reason for bringing up the above examples is to show that there is a limitation with the visualization which we believe auditory display can help to overcome.

In the domain of sonification research, the techniques used in sonifying data have been predominantly circled around Parameter Mapping Sonification, Audification and Auditory Icon/Earcon-based displays. The former technique uses data to drive parameters of a sound generation engine (commonly a synthesizer). While there are many possibilities to create fruitful mappings from data to synthesis parameters, we cannot avoid the sonification researcher's own subjective decision on the mapping and scaling process. In Icon-based displays the two main techniques are Auditory Icons and Earcons, where collective sound pieces or melodies are used to represent certain state of the data vector. These types of auditory displays rely solely on a subjective mapping process. The design bias can be problematic in exploratory data analysis because of the absence of explicit knowledge of the data, e.g. switching the scaling or polarity of the mapping between a data channel and the pitch of synthesizer can lead to very different auditory patterns, prompting to different judgements.

2. LITERATURE OF SONIFICATION FOR EXPLORATORY DATA ANALYSIS

There is a body of work on parameter-mapping sonification for general data inspection[1]. For reasons of limited space we merely mention a few. An early on research of using auditory display in exploratory data analysis can be found in [4], where Flowers and Hauer looked into the efficiency of using auditory display to study data in comparison to histogram and Box-Whisker plots. Peres and Lane also studied the effectiveness of sonification of box plots in [5]. Flowers et al. examined data using auditory and visual display of scatterplots for bivariate data in [6]. The remainder of this section provides pointers to examples of developing general structure-specific sonification models for data analysis rather than targeting a specific type of dataset.

Hermann and Ritter introduced the Model-Based Sonifcation in [7], a paper in which they exemplify the idea with two sonification models: Particle Trajectory Sonification and Data Sonograms. In [8] they introduced a model to sonify data w.r.t. its principal curve. Further sonification models were introcuced in series, such as a local heat exploration model (LHEM) [9], Growing Neural Gas Sonification for exploring intrinsic dimensionality of data and the Markov-chain Monte Carlo Sonification model [10], most of them summarized in [11]. Multi-touch interaction for data sonograms using MBS was carried out in [12]. Crystallization Sonification was introduced in 2002 in [13], this model aims at exploring the intrinsic data dimensionality around user-selected points in relation to the global data dimensionality [14].

[1] throughout the ICAD proceedings

3. TOWARDS REAL-TIME INTERACTION IN PTS

The Particle Trajectory Sonification (PTS) aims at exploring the clustering of vectorial data [15]. This model was initially presented in [7] in 1999. At that time, however, the limitation in computing power did not allow a responsive (real-time) interaction to really treat the vectorial data as a 'virtual physical instrument', onto which the user can apply excitatory interactions to receive auditory feedback instantly in a tightly closed loop. Instead, it took up to minutes for reasonably complex problems to render usable sonifications, interrupting the control loop significantly. In 2016, the computing power has vastly improved. So for the first time, we can achieve (and explore the benefits of) a real-time experience of interaction with such complex sonification models in unprecedented detail and complexity.

The core objective of this research is to optimize the model to speed up in order to sonify multiple particles of larger high-dimensional datasets at a latency that is suitable for real-time interaction. To define a latency threshold to be able to consider as real-time interaction, also take into account that the interactive sonification is a type of sonic interaction. It is natural to assume a similar standard as the latency of a digital musical instrument, especially physical model instruments. In 2002, Wessel and Wright suggested that less than 10ms should be an ideal latency for performing a digital musical instrument [16]. A physical piano however can have even up to 35ms for pp notes [17]. Although it will be an ideal scenario if the MBS can be interacted within such a low latency level, it still seen rather ambitious when sonifying larger datasets. As mentioned in the later section (Sec 5), sonifying large and high-dimensional data may take upto a few seconds. Thus a significant speed up approach is required and also we loosen the definition of real-time interaction to within 300 ms in this particular case as user only needs to receive a relatively quick response for the purpose of analysis of the data rather than performing acute isochronous sequences as would be equipped for musical performance.

We start the presentation in Section 4 with details about the PTS and how this model can be dynamically sonified. Section 5 explores different approaches of implementing the sonification model, followed by a discussion and summary.

4. PARTICLE TRAJECTORY SONIFICATION

4.1. Model Definition

The Particle Trajectory Sonification Model is a Model-Based Sonification to analyze the clustering information of high-dimensional datasets through probing a data-driven potential function by dynamic test particles. Note that this model allows the analysis of multivariate data clustering without requiring to carry out any other clustering analysis beforehand.

The workflow of the model is as follows: a potential function $V(\vec{x})$ is constructed from the given high-dimensional dataset by superimposing data point potential functions which are centered in an Euclidean model space of same dimensionality as the data at each data point's coordinates. The overall potential is just the superimposition of all data point contributions at any arbitrary location in the high-dimensional model space. For sonification, particles with a given initial kinetic energy are injected into the model space. They move around according to a given dynamics (Newton's law plus a friction force). The resulting sonification

is obtained by adding the particles' kinetic energy as function of time.

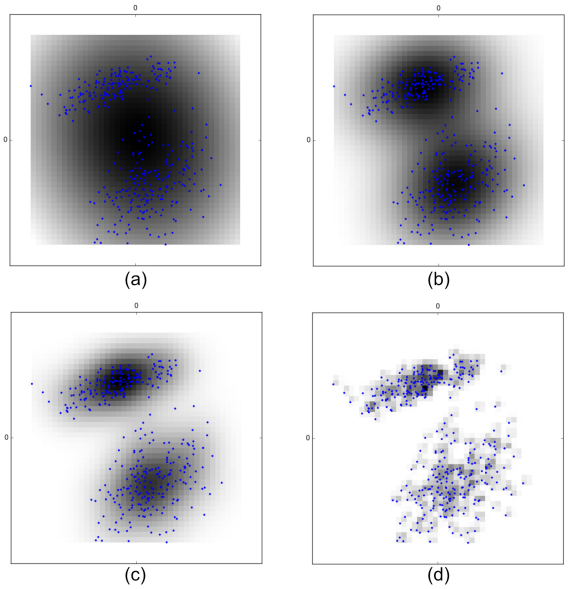


Figure 1: *Plots of the data potential for a 2D toy problem. (a)* $\sigma = 0.4$*; (b)* $\sigma = 0.16$*; (c)* $\sigma = 0.07$*; (d)* $\sigma = 0.01$.

For a formal definition, assume a given data matrix $X \in M(N \times d, \mathbb{R})$ whose row vectors are $\vec{x}_i^T, i = 1, ..., N$. The model space is an Euclidean vector space $\mathbb{R}^d$ in which points are fixed for each data point at coordinates $\vec{x}_i$. Assume an injected particle to have the coordinate $\vec{x}$, then the particle experiences a 'gravitational'[2] potential:

$$V(\vec{x}) = \sum_{i=1}^{N} \Phi(|\vec{x} - \vec{x_i}|) \tag{1}$$

where $\Phi(r)$ is the potential function of a data point defined by:

$$\Phi(r) = -\exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{2}$$

where σ is the interaction length, which determines the resolution of the potential. Fig. 1 shows how σ affects the data potential $V(\vec{x})$ of a two-dimensional data set. When σ^2 is much larger than the average variation over dimensions, V exhibits only a single global minimum near the dataset mean (cf. Fig. 1a). With decreasing σ, local minima may arise corresponding to clusters in the data (cf. Fig. 1b & c). Yet if σ decreases further to smaller values than the average distance between the nearest neighbors of data points, each data points' potential trough is separated, thus we get N local minima (cf. Fig. 1d).

An injected particle of mass m is given a random initial kinetic energy W_{kin} yet so low that the particle can't escape from the data, i.e. $W_{kin} < -V(\vec{x})$. From $W_{kin} = mv^2/2$ we obtain the absolute velocity and set up its random initial velocity vector $\vec{v}$ in d dimensions. Newton's law of motions describe how the particle moves in model space. Numerical integration of the equation of motion with $\Delta t = d_t$ yields the following updates for the particle's position $\vec{x}$ and velocity $\vec{v}$:

$$\vec{v} := r\vec{v} + d_t \frac{-\nabla V}{m} \tag{3}$$

where r is the energy loss ratio due to friction of the model space.

$$\vec{x} := \vec{x} + d_t \vec{v} \tag{4}$$

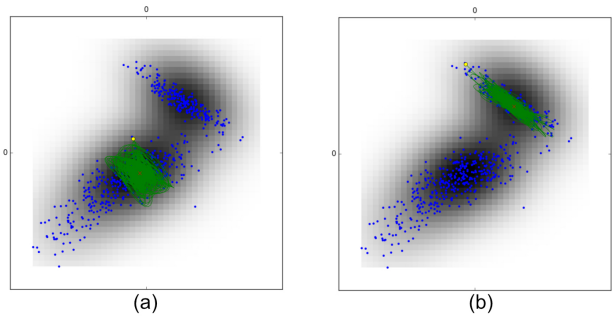


Figure 2: *Examples of the trajectories of two particles are injected in each one of the clusters in* 5000 *iterations. The green lines represent the trajectory. The yellow circle indicates the start coordinates of the particle and the red cross is the end of the trajectory.*

As aforementioned the sonification is the time series of kinetic energy[3]. The frequency and spectral complexity of the sound depend on rate of change of the velocity v. This means two key properties of the sonification in order for user to listen and understand the data:

- Particles attracted to a larger cluster exhibit a higher pitch than those attracted to a small cluster due to the stronger gravitational pull.
- If a particle started at the edge of a cluster, as it oscillates through the mode and (by friction) converges slowly to the cluster center (i.e. the mode), the spectral complexity of the sound decreases until eventually a sine wave like tone. This is due to the particle becoming less and less affected by the non-harmonic shape given by the tails of the potential function V .

An example of the particle movement can be seen in Fig. 2. Two particles were injected into different clusters (the yellow dots are the initial locations). After the initialization, the particles started moving in model space. They are pulled by the collective attractive forces of all model elements and due to friction they converge (in this case) to the local minimums of the clusters the particle belonged. Subject to different sizes of the clusters, their potential functions are different causing different velocity changes throughout the trajectories.

From the perspective of the time series of the particle's kinetic energy, the sonification is structurally merely the audification[4] [1] of the kinetic energy as a function of time.

[2] gravity works different in this model space than in our universe

[3] Equivalent to the square velocity vector since mass is a constant

[4] Audification is a technique that considers the data vector directly as an audio vector and plays the vector back in audio rate with downsampling or upsampling if required.

A video example of the sonification can be found in Sec. 7. Please refer to 'PTS-parameters.mp4' for a demonstration how σ and r affect the sound.

4.2. Exploring example data using the system

In Fig. 3, an example is presented of using the system to distinguish two separate clusters. In this example, we used a mixture of Gaussians to create data with the controlled features, here 4 clusters, $N = 980, d = 6$. The figure suggests that the right part of the scatter plot is a single cluster. However, the trajectories show different convergence targets. Fig. 3a & 3b each demonstrates the trajectory of particles attracted by different clusters. The particles converged to two different areas, leading to two different textures in sound as also rendered visible in the spectrogram respectively. This indicates that there are indeed two clusters whose discrimination is visually impossible as they overlap in the scatter plot.

Please refer to video file 'PTS-clusters.mp4' for a demonstration on how to use the sonification model for detecting different clusters, for finding cluster's edges/center and outliers.

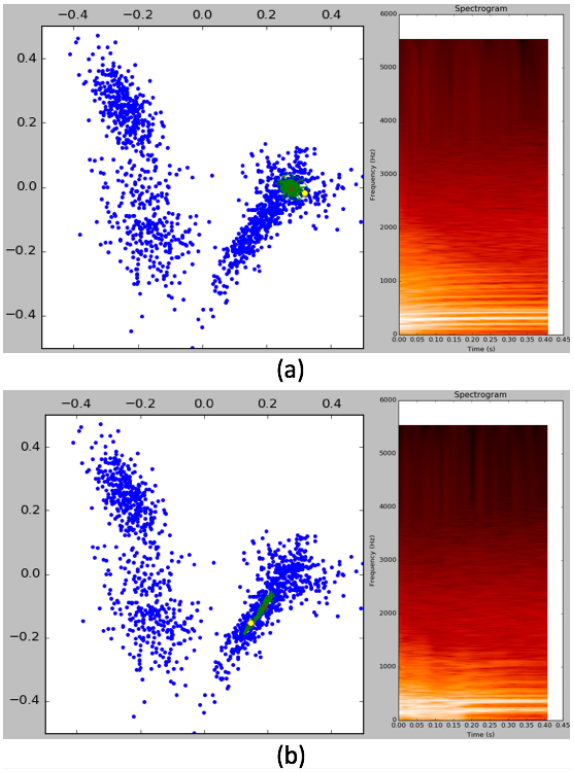


Figure 3: *Comparison of the two particle trajectories and their spectrograms.*

5. MODEL OPTIMIZATIONS WITH VECTOR QUANTIZATION AND PARALLEL COMPUTING

The previous section presented the mathematical model and sound examples for injecting a single particle. The analysis of clustering is about understanding the grouping of data points, which is a zonal property. Thus it will be more meaningful to inject multiple particles simultaneously and listen to their collective sound in a particular zone. The result sonification can either be listened individually or holistically by superimposing all particles' energy and then performing a normalization.

The following section presents multiple approaches of optimizing the dataset with vector quantization, as well as parallelism with OpenCL framework [18].

5.1. Baseline method

The baseline method is written is Python using the C-extension library Cython [19] for calculating the particle trajectory. The squared velocity vector is played back as sound using pyo's FIFO player [20] module. Using the FIFO player, the particle trajectory can be rendered in smaller chunks (per buffer size N_s) continuously that are then pushed to the queue of the FIFO player for continuous playback. A smaller buffer size can lower the latency as long as the rendering time t is below $t < N_s/f_s$, where f_s is the sampling rate. For PTS, we choose a sampling rate of $f_s = 11025$ because there is very little audible information in the sonification at higher frequencies. We use a buffer size $N_s = 256$. This implementation is originally targeted for single PTS.

In terms of the trigger latency, we define 300 ms as the threshold below which interaction (i.e. insert particles and hear back the sound) can be regarded as real-time.

As for the performance, the computational cost is proportional to the number of particles, which was set between 1 to 100. As shown in Fig. 4, for smaller datasets (200×5), Cython implementation can still achieve satisfactory latency for real-time interaction. As for the 2000×5 dataset, rendering multiple-particle trajectories slowed down significantly, and it became unsuitable above 30 particles. This results prompts to a requirement for significant speedup.

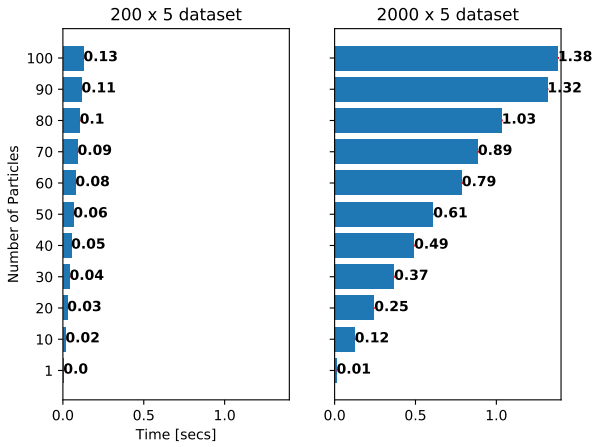


Figure 4: *Performance for different numbers of particles using Cython without parallelization.*

5.2. Vector Quantization

When dealing with larger datasets, e.g. $N > 10^3$ data points, vector quantization allows to reduce the resolution while still maintaining the general clustering structure. Here, we apply the k-means vector quantization to reduce the size of a given dataset into k prototypes minimizing the cost function

$$E = \sum_{i=1}^{N} \sum_{j=1}^{k} h_{ij} \left\| \vec{x}_i - \vec{v}_j \right\|^2 \quad (5)$$

where $h_{ij} = 1$ if $\vec{x}_i$ is nearest to prototype $\vec{v}_j$, else $h_{ij} = 0$. Note that the number of prototypes is still as high as possible, e.g. around 1000, and not selected according to the number of expected clusters. A detailed analysis as to how data set reduction affects the PTS will be published elsewhere. Fig. 5 shows a linear change (note that both x and y-axis are log-scaled) in the computational time of the PTS algorithm with 1 particle only after applying different levels of vector quantizations to a $10^4 \times 5$ dataset. Based on

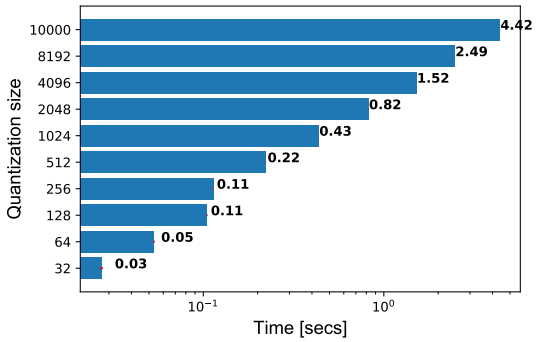


Figure 5: *Benchmark of the vector quantization of a $10^4 \times 5$ normalized dataset. The test was run on a Windows desktop with Intel i5-3350 2.7GHz CPU.*

Fig. 5, in order to have a responsive interaction for a single particle (i.e. latency < 0.3 s), the total numbers $N \times d$ of a data matrix should be approximately < 15000. Thus, given a new data matrix $X \in N \times d, \mathbb{R})$, if the total number of cells is greater than the threshold, a mild vector quantization is of k prototypes is applied, e.g. with $k = \text{round}(15000/d)$.

The vector quantization can be precompiled prior to interaction and sonification as a data reduction method, thus the compute time for the quantization does not affect the time for the sonification.

5.3. Parallel Computing with OpenCL

We paralleled the computation of multiple particles' sonification using OpenCL. The OpenCL framework allows parallelization to be executed on either CPUs or GPUs. In both cases, the N_p particles are distributed to the parallel kernels. Each kernel shares the same data matrix and model parameters, the only difference is the initial location of the particle. As a result, each kernel returns the trajectory and kinetic energy vector for the assigned particle. The performance results are presented in the next subsection.

5.4. Performance of Different Process Units

We tested the new model implementation on 1 CPU (Intel(R) Core(TM) i5-5287U CPU @ 2.90 GHz, dual- core) and 2 GPU Nvidia GeForce 940M and Nvidia GeForce GTX 970). These three units cover a common range of consumer processing units. Due to the limitation in equipment, the latest generation graphic cards in 2017 such as GeForce GTX 1000 series are not tested. However, our selection shall provide an insight into the performance of the PTS model using common computers.

Two random datasets are selected for the benchmarks. A smaller data matrix $X_1 \in M(200 \times 5, \mathbb{R})$ and a larger data matrix $X_2 \in M(2000 \times 5, \mathbb{R})$. Notice that the total number of cells of X_2 equals the threshold for applying vector quantization, hence vector quantization was not applied.

Comparing the baseline result from Fig. 4 with using OpenCL with CPU (Fig. 6), the latter achieved a mean acceleration ratio of 1.67 in X_1 and 1.59 in X_2 for multiple particles ($N_p > 1$).

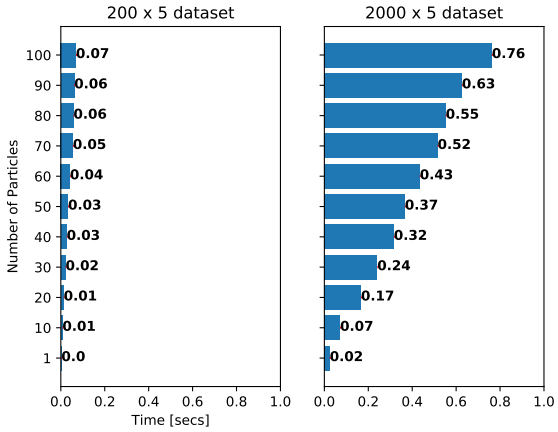


Figure 6: *Performance of PTS written in OpenCL using the Intel(R) Core(TM) i5-5287U CPU @ 2.90GHz.*

When testing with the GPUs, Fig. 7 & 8 show that there is a constant offset in the computation time, which is due to the data being copied in and out of OpenCL kernel at each call. The offset time is proportional to the dataset size, thus for the larger dataset X_2 the latency is over threshold. On the other hand, since the OpenCL implementation paralleled each particle, increasing the number of particles does not lead to a proportional increase in the computing time. Further tests with a higher number of particles $N_p > 10^3$ showed that compared to CPU we can gain a 30-fold speedup before compute time starts to increase linearly also with GPU.

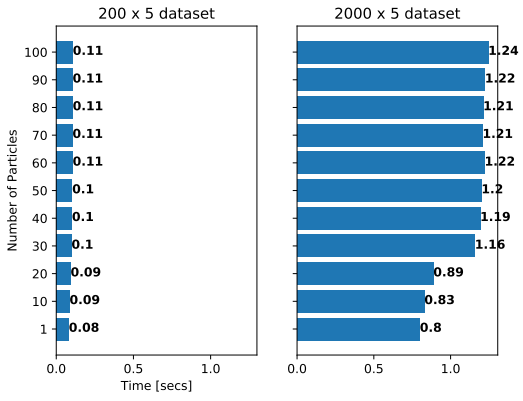


Figure 7: *Performance of PTS written in OpenCL using the Nividia GeForce GTX 970 graphic card.*

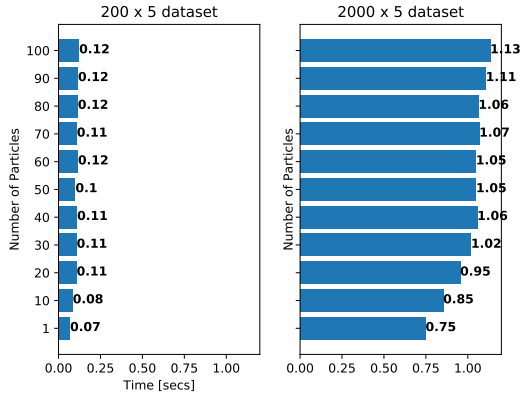


Figure 8: *Performance of PTS written in OpenCL using the Nividia GeForce 940m graphic card.*

5.5. Hybrid parallelism between CPU and GPU

Section 5.1 – 5.4 led to the following conclusions:

1. The Cython implementation is 'in serial' thus the computation time increased linearly as N_p increased.
2. Comparing this to using Cython and OpenCL on Intel i5 CPU, the OpenCL implementation is faster. But it is not efficient at dealing with larger amount of particles.
3. Implementing OpenCL on GPU has a strong ability for paralleling large amount of particles but currently prolonged by the data I/O overhead (offset) between memory and graphic card units.

Based on these 3 results, we propose a hybrid OpenCL solution that utilizes the CPU for smaller N_p but then automatically switches to GPU-based rendition if N_p becomes larger. The variation in N_p will depend on the user's interactions, e.g. tapping an area with finger or palm. The threshold may vary due to the computer's processing power, yet it can be pre-calculated when a new dataset is loaded.

We conducted a test to study how the following three key variables that affect the computational time:

- $N_D = N \times d$ is the total size of the dataset. This variable affects is the main factor for the data input overhead when using GPU. It also contributes the cost of the potential function (cf. eq. (1)).
- N_v is the trajectory vector (aka. the audio buffer size), which affects the overheads of both input and output (passing the vector from GPU back to memory).
- N_p is the number of particles, which only affects the trajectory's computation cost.

From our benchmarks we can model the relationships between the three aforementioned variables and computational time. For GPU we expect:

$$t = aN_DN_v + bN_v + \begin{cases} cN_DN_v & \text{if } N_p <= N_p^o \\ c\frac{N_p}{N_p^o}N_DN_v & \text{else} \end{cases} \quad (6)$$

whereas for CPU, we expect

$$t = dN_DN_vN_p \ , \quad (7)$$

where a, b, c, d are coefficients and N_p^o is the threshold of maximum GPU capacity for calculation in parallel at at time. Eq. (6) addresses that the computation cost is based on three parts: input offset, output offset and cost for rendering the sonification and trajectory. However, the cost does not increase when $N_p \leq N_p^o$. Using OpenCL with CPU does not suffer from the copying offsets but only lacks the benefit of allowing a larger number of particles to run without affecting the time.

We then tested the assumption using the GTX970 graphic card and the Intel i5 CPU with different N_D, N_v and N_p and their correspondent computational time t (cf. Fig. 9). However, instead of plotting t as function of N_P, we depict $t/(N_D \cdot N_v)$, which gives the unit computational time that is only relevant to the number of particles plotted on the x-axis. In Fig. 9, each line (apart from the black straight line) represents a specific combination of N_D and N_v. The black straight line is the linear regression of the CPU's OpenCL performance[5].

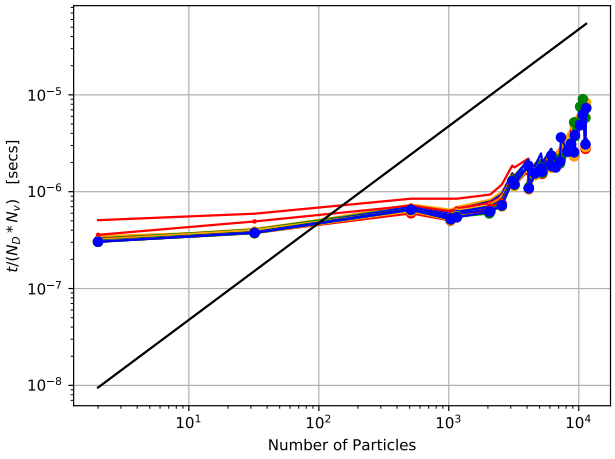


Figure 9: *Normalized computation time ($t/(N_DN_v)$) comparison. The markersizes of the data points represents is mapped to N_D. Color coding is mapped to N_v.*

The figure shows that the approximate maximum number of particles for which parallel computing most significantly increases performance (N_p^o is peaked at 2048, from then on the required time increased linearly with N_p. When $N_p < N_p^o$, the computation cost stayed relatively flat with minor increase as N_p increased, which is due to $N_p \times N_v$ of kinetic energy vector (for sonification) and $N_p \times N_v \times d$ of position matrix (for visualization) to be copied from GPU back to the memory. Also, the offset of the first data point ($N_p = 1$) indicates the input offset. At a larger number of particles ($N_p > 100$), GPU shows superiority over CPU and the speedup exhibits a maximum at $N_p^o \approx 2048$, which is about 30 times faster than CPU.

In result, for the tested hardware configuration we arrive at the following conclusions:

1. It is more effective to use OpenCL with CPU than GPU when $N_p < 100$, vice versa.

[5]The high linearity we found when using CPU at lower particle number (< 100) led us to use linear regression line rather than the actual computational cost.

2. When $100 < N_p < 2048$, it is more efficient to use GPU to better exploit the potential of parrallelism.
3. When $N_p > 2048$, the speedup is maximized with a factor of ≈ 30.
4. A significant overhead persists for our current approach due to data being copied in and out of the GPU.

We then put the hybrid method into test and set the switching threshold to 100. In this test (cf. Fig. 10), we tested different numbers of particles and data sizes. As mentioned before, 300 ms is the latency threshold in order to be considered suitable for real-time interaction. Based on the graph, most of the tested data sizes are suitable for real-time interaction when N_p is smaller than 100. For $N_D < 1400$, low latency can still be still achieved for higher N_p values. But as $N_D > 1400$ increased, the latency is greater than 300 ms for larger numbers of particles. However, this issue can be addressed by using the above-proposed vector quantization approach presented in Sec. 5.2.

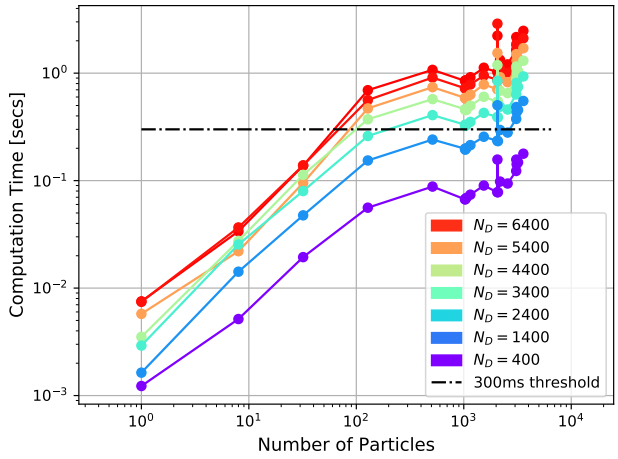


Figure 10: *Plot of the computation time t with the hybrid method against N_p. The dashed line indicates our defined threshold for real-time interactivity.*

6. DISCUSSION & CONCLUSION

This paper continues the research on Model-Based Sonification. For the example of Particle Trajectory Sonification we implemented data optimization and parallel computing procedures to increase the responsiveness when interacting with high-dimensional complex data. The PTS model was initially designed to analyze clustering features of usually high-dimensional datasets offline as the computational time may take up to minutes and even longer. In this project, we proposed several methods to increase the total performance when sonifying multiple particles' trajectories for larger datasets. This could pave the way towards an interactive structure-specific[6] toolbox for the exploratory analysis of high-dimensional data. Firstly, vector quantization can effectively set up a time cap of 300 ms for a single particle trajectory while maintaining the data structure and thus enable PTS for larger datasets. Secondly, we observed that the OpenCL implementation via CPU provides a moderate speedup compared to the baseline Cython implementation. With a higher amount of GPU process units, increasing the amount of particles does not increase the time proportionally. At our test, we found a maximum of 30 times faster computation as compared to CPU for larger number of particles (> 2000). However, the data I/O overhead between memory and the graphics card is still large for our current implementation. For the tested computer configuration, the hybrid OpenCL implementation can ensure low latency ($<$300 ms) for $N_p < 100$ even with large dataset, but it is found to be difficult to keep the latency low as the number of particles N_p becomes very large.

For our next steps, we plan to reimplement our OpenCL algorithm in hope to eliminate input offsets of GPU by allowing the data matrix to remain stored in the graphic units instead of passing it at each call. Also, we currently work on a distance-matrix-based method for choosing the most influential neighbors relative to the current particle's position at each iteration and discarding the rest. This could lead to another significant speed up in combination with vector quantization. These together could potentially lead to a significant speedup allowing thousands of particles to be sonified in high-dimensional datasets in real-time.

The advantage of listening to the clustering information via PTS over the visualization of the potential map (cf. Fig. 1, 2), is that potential maps can only be computed for rather low-dimensional problems (< 3 dimensions), whereas with few simple sound probes, the user can navigate the full potential function and quickly explore prevalent potential troughs (corresponding to clusters) at different scales of resolution.

[6] in contrast to data-specific

7. LINK

Supplementary material for this paper (media files) are available via the DOI: `10.4119/unibi/2911345`

8. ACKNOWLEDGMENT

This research was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG). We wish to thank our student worker Akhil Jain for programming support.

9. REFERENCES

[1] T. Hermann, A. Hunt, and J. G. Neuhoff, Eds., *The Sonification Handbook*. Logos Verlag, 2011.

[2] A. Inselberg and B. Dimsdale, "Parallel coordinates: A tool for visualizing multidimensional geometry," *Proc. IEEE Visualization*, pp. 361–378, 1999.

[3] M. Graham and J. Kennedy, "Using curves to enhance parallel coordinate visualisations," *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on*, pp. 10–16, July 2003.

[4] J. H. Flowers and T. A. Hauer, ""sound" alternatives to visual graphics for exploratory data analysis," *Behavior Research Methods, Instruments, & Computers*, vol. 25, no. 2, pp. 242–249, 1993.

[5] S. C. Peres and D. M. Lane, "Sonifcation of statistical graphs," *Proceedings of the 2003 International Conference on Auditory Display*, 2003.

[6] J. H. Flowers, D. C. Buhman, and K. D. Turnage, "Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 39, pp. 341–351, 1997.

[7] T. Hermann and H. Ritter, "Listen to your data: Model-based sonification for data analysis," *Advances in intelligent computing and multimedia systems*, vol. 8, pp. 189–194, 1999.

[8] T. Hermann, P. Meinicke, and H. Ritter, "Principle curve sonification," 2000.

[9] T. Bovermann, T. Hermann, and R. Helge, "The local heat exploration model for interactive sonification," *International Conference on Auditory Display*, 2005.

[10] T. Hermann, M. H. Hansen, and H. Ritter, "Sonification of markov chain monte carlo simulations," *International Conference on Auditory Display*, 2001.

[11] T. Hermann, *The Sonifcation Handbook*. Logos Verlag, 2011, ch. Model-Based Sonification, pp. 399–425.

[12] R. Tünnermann and T. Hermann, "Multi-touch interactions for model-based sonifcation," *Proceedings of the 15th International Conference on Auditory Display*, 2009.

[13] T. Hermann and H. Ritter, "Crystallization sonifcation of high-dimensional datasets," *Proceedings of the 2002 International Conference on Auditory Display*, 2002.

[14] T. Hermann, "Sonification for exploratory data analysis," Ph.D. dissertation, Bielefeld University, 2002.

[15] T. Hermann and H. Ritter, "Model-based sonification revisited—authors' comments on hermann and ritter, icad 2002," *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 4, pp. 559–563, 2005.

[16] D. Wessel and M. Wright, "Problems and prospects for intimate musical control of computers," *Computer Music Journal*, vol. 26, no. 3, pp. 11–22, 2002.

[17] A. Anders and E. Jansson, "From touch to string vibrations - the iniitla course of the piano tone," *The Journal of the Acoustical Society of America*, vol. 81, no. S1, pp. S61–S61, 1987.

[18] J. E. Stone, D. Gohara, and G. Shi, "Opencl: A parallel programming standard for heterogeneous computing systems," *Computing in Science & Engineering*, vol. 12, no. 1-2, pp. 66–73, 2010.

[19] S. Behnel, R. Bradshaw, D. S. Sljebotn, and G. Ewing, "Cython: C-exsions for python," 2008. [Online]. Available: http://cython.org

[20] O. Bélanger, "Pyo," 2012. [Online]. Available: http://ajaxsoundstudio.com/team/

AN IMMERSIVE VIRTUAL ENVIRONMENT FOR CONGRUENT AUDIO-VISUAL SPATIALIZED DATA SONIFICATIONS

Samuel Chabot and Jonas Braasch

Graduate Program in Architectural Acoustics
Rensselaer Polytechnic Institute
Greene Building, 110 8th St,
Troy, NY 12180, USA
chabos2@rpi.edu

ABSTRACT

The use of spatialization techniques in data sonification provides system designers with an additional tool for conveying information to users. Oftentimes, spatialized data sets are meant to be experienced by a single or few users at a time. Projects at Rensselaers Collaborative-Research Augmented Immersive Virtual Environment Laboratory allow even large groups of collaborators to work within a shared virtual environment system. The lab provides an equal emphasis on the visual and audio system, with a nearly 360° panoramic display and 128-loudspeaker array housed behind the acoustically-transparent screen. The space allows for dynamic switching between immersions in recreations of physical scenes and presentations of abstract or symbolic data. Content creation for the space is not a complex process-the entire display is essentially a single desktop and straight-forward tools such as the Virtual Microphone Control allow for dynamic real-time spatialization. With the ability to target individual channels in the array, audio-visual congruency is achieved. The loudspeaker array creates a high-spatial density soundfield within which users are able to freely explore due to the virtual elimination of a so-called "sweet-spot."

1. INTRODUCTION

In the inherently interdisciplinary field of sonification, the universal definition is the use of non-speech audio to convey information [1]. Due to a number of key advantages presented by the human auditory system, the topic continues to expand into a largely explored field of research. These advantages include a strong sense for pattern recognition and an ability to perceive subtle and transient changes that might otherwise go overlooked or unrecognized. Additionally, humans are constantly and rapidly decoding complex auditory scenes presented by their environment [2]. Benefits such as these make sonification a powerful tool in the area of data analysis.

In recent years, multiple use cases have emerged that map qualities of a set of data to specific acoustical properties of produced sound to convey information. This method of matching characteristics of data sets to acoustical details is well-known as a parameter-mapping technique. Parameter-mapping is a popular framework for data sonification because it makes use of sound's multidimensionality to convey changes and trends in data. Qualities of the produced sound, such as pitch and timbre, rhythm and tempo, and loudness, to name a few, are correlated to characteristics of the data [3].

Another function of a sonification system is the ability to harness spatialization of the produced sound. By doing so, the system alters the perceived location of auditory streams, and consequently employs the human ear's capacity for attending to multiple audio cues simultaneously. Strategic positioning can strongly influence the conveyance of information and immersion within the data. There are a number of techniques to create spatialized sound, each with benefits and limitations.

2. SPATIALIZATION TECHNIQUES

Vector Based Amplitude Panning (VBAP) is a popular method for spatializing sound. The technique makes use of pairs or triplets of loudspeakers who's individual gains are controlled such that the perception of virtual sound sources is created within the space between [4]. The result is effective at conveying the general direction of an incoming sound source while remaining computationally tractable. The so-called sweet-spot is flexible enough that a user is not confined to one specific static location. Another favored spatialization technique involves the use of headphones to create the spatialization. Head-Related Transfer Functions (HRTFs) are convolved with the audio signal, most often provided by a general HRTF library. While an individual's own HRTF varies from person to person, the result is a fairly realistic spatialization. However, it can be computationally expensive and is limited to a single simultaneous user. The soundfield is also almost always head-locked such that head movement does not correlate to movement within the soundfield.

A powerful strategy with continuously growing interest is the Wave-Field Synthesis technique. By using a dense array of loudspeakers, artificial wavefronts from virtual sound sources can be reproduced. The signal produced at each speaker is appropriately weighted and delayed such that the synthesis of the entire array creates a discretized reproduction of the desired sound wave [5]. The prime functions of WFS are the high precision with which users perceive both the angle and distance of virtual sound sources, as well as the elimination of the sweet-spot. Unlike an HRTFs head-locked limitation, and in addition to a VBAP method allowing head movement, WFS supports multiple users moving about

https://doi.org/10.21785/icad2017.072

the soundfield while completely retaining an accurate rendering of the sources.

The environments in which these spatial sonifications are deployed play crucial roles in the user experience of the information. Virtual environments have become the forefront of experiential spaces. The virtualization of the world around has been discussed and developed for a number of decades-in the middle of the twentieth century, researchers were already creating what would be the fundamental basis of the virtual environments of today. Contemporary projects attempt to transport the user(s) to environments that may not otherwise be immediately accessible. An example of such a space is the Allosphere at the University of California at Santa Barbara, which incorporates a large surrounding screen and complex audio systems for the purpose of immersing the user(s) [6]. Rensselaer has recently developed such a space to enhance and build upon these foundations.

3. COLLABORATIVE-RESEARCH AUGMENTED IMMERSIVE VIRTUAL ENVIRONMENT LABORATORY

The Collaborative-Research Augmented Immersive Virtual Environment Laboratory (CRAIVE-Lab) is a state-of-the-art immersive virtual environment developed at Rensselaer Polytechnic Institute. The goal of this space is to support a multi-modal workspace which puts equivalent emphasis on both the auditory display and visual display. The lab is intended for use not only by a single user but a large group of users simultaneously, with a focus on both scientific and artistic works, and often the gray area in between. To accommodate a large group of people, the workspace is sized accordingly, measuring 10 m by 12 m.

3.1. Video rendering

The space is outfitted with a 4.3 m tall, nearly 360° projector screen surrounding the perimeter of the lab. The screen is rectangular in shape, with rounded corners between the four sides. A custom brace with C-clamps holds the screen taut smooth through the corners, shown in the right image of figure 2.

Creating a single continuous display across the entire screen requires the use of eight high-performance projectors. To allow users the ability to approach the screen, short-throw projectors were chosen. Users can stand as close as nearly a meter in front of the screen before casting any shadow on the display. Each projector runs at a resolution of 1920x1200, creating a single cohesive and continuous desktop of 15360x1200 pixels. A Dell workstation housing two NVIDIA Quadro K5200 cards with four video outputs each allows the system to render across all eight projectors with one computer.

The eight projectors and screen were meticulously aligned using laser levels and calibrated using a software tool by Pixelwix, which provided 8000 reference points at pixel resolution. The left image of Fig. 2 shows the Pixelwix calibration grid for one calibrated corner. This program is also able to stitch the projectors into a single screen, correcting for the room's geometry and smoothly blending the individual projectors together.

The projection screen itself is made of a microperforated PVC material, rendering it acoustically transparent. This allows for sound transmission, critical because located behind the screen are 128 loudspeakers for sonification purposes.

Since the visual component is essentially a single desktop display that is already blended and corrected for the rooms geometry, content-creation is a straight-forward process. Visualizations need only be created for the entire span of the 15k pixel width of the screen. For many programs such as Max/MSP/Jitter or Photoshop, this only requires defining the workspace with the correct pixel amounts. High-quality replications of real-life locations can be created using software for stitching together photographs into 360°panoramas. Recently, the use of a 360°recording device, the Freedom360 GoPro rig, has allowed for the reproduction of immersive video for the lab as well. Renderings of artificial spaces and symbolic visualizations allow users to stand in environments that may otherwise not be possible to. Figure 1 shows a stitched panoramic rendering of a Roman amphitheater found in Germany.

3.2. Audio rendering

The laboratory is outfitted with a total of 134 loudspeakers, 128 of which are located in a horizontal array around the perimeter of the space approximately at ear-height behind the projector screen. Six of the loudspeakers are mounted from the ceiling and directed downward into the space to assist in further enhancing audio immersion. A heavy acoustic curtain, which spans the entire height and length of the screen, is hung behind the horizontal array to provide strong damping of the physical room's response. Carpeting will be installed soon over the current hard-surfaced floor to provide additional dampening.

All loudspeakers are individually routed using XLR cabling (totaling over 2 miles!) to a digital audio workstation (DAW) where they are connected in consecutive sets to 16 eight-channel preamplifiers/optical interfaces. A dedicated Mac Pro equipped with an RME HDSPe MADI FX sound card controls the audio output. The computer uses the sound card to send all 128 channels of audio for the horizontal array over two optical MADI outputs. These MADI outputs are each received by an eight-channel format converter, split into eight channels of AES-formatted output, and, subsequently, received by the preamplifiers for output to the loudspeakers. This hardware setup allows each loudspeaker channel to be accessed and targeted individually. The additional six hanging speakers are fed through a separate M-Audio 1814 interface because all of the available channels of the format converters and preamplifiers are occupied by the horizontal array. Figure 6 provides a helpful diagram, which shows the dimensions and loudspeaker channel layout of the space, courtesy of J. Carter of Rensselear. Additionally, a Sennheiser wireless microphone system with four microphone units is installed for audio input to the system. Figure 5 provides a visual of the workstations and audio interfaces that support the laboratory

The system is flexible in its production of sound, and multiple approaches can be taken. Because each channel can be addressed individually, many programs are able to simply specify which loudspeaker channel(s) to target. This allows a sound designer the ability to quickly and easily distinguish between global and local sounds. Output of auditory streams can be connected to all channels, thereby creating global sounds in the space. Likewise, these outputs may be concentrated in a single or clustered groups of channels creating local sounds in targeted regions of the lab. Combinations of these two concepts create soundfields in which the opportunity to move about the space becomes almost critical to exploring and discerning all possible information.

Because of the high density nature of the horizontal loudspeaker array, spatialization techniques that harness wave-field synthesis fundamentals are effective ways of creating soundfields

Figure 1: Panoramic rendering of Roman amphitheather in Germany

in the lab. A powerful tool developed by Jonas Braasch is the Virtual Microphone Control (ViMiC) program implemented within the Max/MSP visual coding environment. ViMiC is designed for the flexible, real-time spatialization of sound sources [7]. The program creates a computer-generated virtual space within which are placed virtual sound sources and receivers. These sound sources are user-defined audio inputs to the ViMiC module and the receivers are virtual microphones. The system calculates what would be received by the virtual microphones within the virtual space when the virtual sources radiate their sound. The appropriate pressure and delay at each microphone is considered. The received sound is then mapped back to reality through corresponding loudspeakers. The system contains many variables that the user is able to define. These include the microphone directivity patterns, sound source radiation patterns, strength of distance attenuation, and source and receiver locations and orientations within the virtual environment. Another useful ability of the Virtual Microphone Control is its implementation of the OpenSound Control protocol [8].

For use in the laboratory, the Virtual Microphone Control can be used to virtually deploy microphones around the perimeter of the workspace. Sound sources then placed within the virtual space are received by the array of virtual microphones and mapped to reality over the loudspeaker array of the CRAIVE-Lab. Fig 4 shows an example of a sound source radiating within the virtual space, and the corresponding wavefront that is reproduced at the loudspeaker array. As can also be seen in the figure, a critical benefit of sound reproduction in the CRAIVE-Lab is the elimination of the sweet-spot.

3.3. Other systems

To further enhance the capabilities of the CRAIVE lab, we are currently extending the CRAIVE-Lab with a smart lighting equipment based on six ETC D60 LED fixtures that can vary spatially, temporally, and spectrally, depending on the content being displayed using the eight video projectors to enhance the performance of the users in the CRAIVE-Lab.

An intelligent position-tracking system estimates current user locations and head orientations as well as positioning data for other objects. For the tracking system, a hybrid visual/acoustic sensor system is being used to emulate the humans ability to extract robust information by relying simultaneously on different modalities. A network of six cameras has been installed in the CRAIVE-Lab accompanied by a 16-channel spherical ambisonic microphone with additional peripheral microphones.

Meas. Pos.	1 (L29)	2 (L30)	3 (L31)	4 (L32)	5 (L33)	6 (L34)	7 (L35)
Level	26	30	30	34	31	30	29
ILD	–14.8	–14.2	–11.5	–8.4	–1.9	4.3	8.1
ITD	–0.50	–0.38	-0.27	–0.17	–0.02	0.13	0.23

Table 1: Binaural manikin measurement data or the close measurement position for different loudspeakers (1 m distance between the binaural manikin and the front loudspeaker). From top to bottom: Relative Sound Pressure Levels in decibels, Interaural level differences in decibels, interaural time differences in milliseconds.

4. MEASUREMENTS

4.1. Methods

To examine the spatial abilities of the space, measurements were taken in the CRAIVE-Lab for three different listener positions. All measurement positions are shown in Fig. 6. For the close distance case, a binaural manikin (Neumann K100 [9]) was placed at a 1 meter distance in front of loudspeaker 32 (L32) and the 7 loud-

Figure 2: CRAIVE-Lab Panorama Screen. Left: Pixelwix calibration software. Right: Smooth corner solution of the CRAIVE screen.

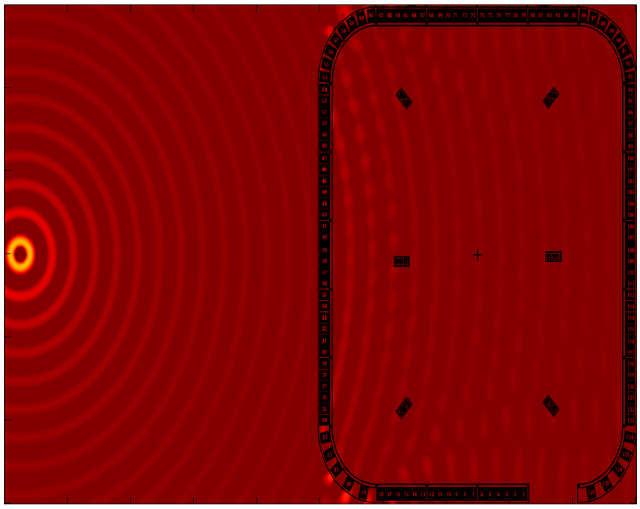


Figure 3: Synthesis of discrete loudspeakers into single wavefront for reproduction of global sound source

Meas. Pos.	1 (L27)	2 (L28)	3 (L29)	4 (L30)	5 (L31)	6 (L32)	7 (L33)
Level	23	24	25	26	26	26	26
ILD	–11.6	–10.9	–10.3	–9.1	–7.2	– 4.4	–1.3
ITD	–0.33	–0.29	–0.25	–0.19	–0.13	–0.06	0.00
Meas. Pos.	8 (L34)	9 (L35)	10 (L36)	11 (L37)			
Level	25	25	24	24			
ILD	1.5	4.3	6.5	7.5			
ITD	0.06	0.13	0.19	0.23			

Table 2: Binaural manikin measurement data or the middle measurement position for different loudspeakers (2.1 m distance between the binaural manikin and the front loudspeaker). From top to bottom: Relative Sound Pressure Levels in decibels, Interaural level differences in decibels, interaural time differences in milliseconds.

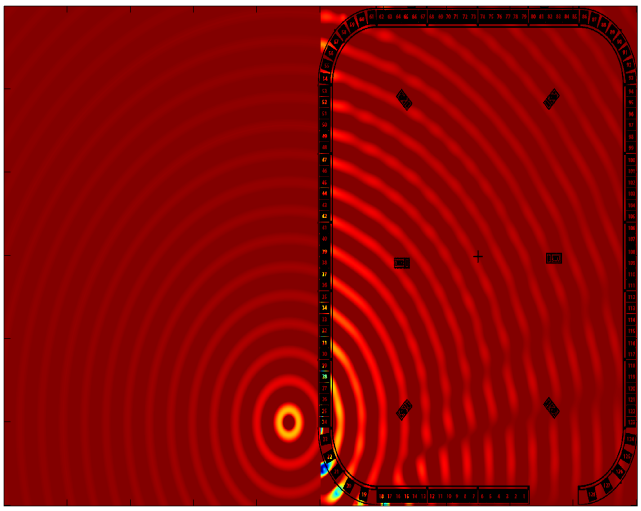


Figure 4: Synthesis of discrete loudspeakers into single wavefront for reproduction of local sound source

Meas. Pos.	1 (L24)	2 (L27)	3 (L30)	4 (L33)	5 (L36)	6 (L39)	7 (L42)
Level	18	19	21	21	21	21	21
ILD	–8.3	–9.1	–8.4	–7.4	–5.6	–2.2	1.3
ITD	–0.38	–0.33	–0.27	–0.21	–0.08	–0.02	0.08
Meas. Pos.	8 (L45)	9 (L48)	10 (L51)	11 (L54)			
Level	19	18	18	18			
ILD	3.6	4.9	5.2	5.4			
ITD	0.15	0.23	0.27	0.33			

Table 3: Binaural manikin measurement data or the far measurement position for different loudspeakers (5 m distance between the binaural manikin and the front loudspeaker). From top to bottom: Relative Sound Pressure Levels in decibels, Interaural level differences in decibels, interaural time differences in milliseconds.

Figure 5: Left and Center: CRAIVE audio rack, right: CRAIVE cabeling.

speaker position were tested (L29–L35). For the mid position, the binaural manikin was placed at a distance of 2 m in front of L32 and the loudspeakers L27 to L37 were measured. For the far position, the binaural manikin was placed in the center of the CRAIVE-Lab (5 m distance from the loudspeaker). A Gaussian noise burst (30 second duration) was used as the test signal, which was measured using the binaural manikin. Relative broadband sound pressure levels were calculated for each position as an average for the left and right ear signals. Interaural level differences (ILD) and interaural time differences (ITDs) were also measured for the high frequency range for ILDs (2–20 kHz) and low frequency range for ITDs (100-2000 Hz).

4.2. Results and Discussion

As expected the loudspeaker right in front of the listener (binaural manikin) has the highest measured level for each of the three binaural-manikin positions – see Tables 1–3. Levels of 34 dB SPL for the close position (L32, 1 m distance from speaker), 26 dB SPL for the mid position (L32, 2 m distance from speaker), 21 dB SPL for the far position (L36, 5 m distance from speaker) are measured. The level reduction can be easily explained by the inverse square law.

Now to look into the approach of local vs. global acoustic viewpoints: as stated previously, this concept builds on the idea that the listener can focus much better on a local speaker to extract its auditory information, because the information to the speaker's sides will roll off much quicker than would be the case if the listener were to stand far away from the speaker. The experimental results reflect the following: at the close position, the loudspeaker levels roll off much quicker with distance from the front position than is the case for the mid or far positions. In the close position the levels of the loudspeakers 3 units away from the front speaker (L32: 34 dB) roll off by approximately 6 dB (L29: 26 dB and L25: 29 dB). For the mid position, the level reduction for the same speakers is only 1 dB, and for the far position, hardly any level reduction was found when the sound was moved three speakers to the right or left (all three measurements, L30, L36 and L42 show a dB level of 21 dB.

4.3. Extending concepts to include wave field synthesis

Included in the next step are the simulation of virtual sound sources to project acoustical information. Using wave field synthesis, it is possible to place a sounds at any point behind (or, with certain restriction, in front of) the screen. At a far distance, a broad wave front will be radiated using multiple speakers that does not roll off to the side – see Fig. 4. Using multiple sound sources place at different distance, local and global audio streams can be created. The global streams are placed in the background so they can be heard from any position in the lab. The local streams are presented from local loudspeakers. For the latter, the listener can move closer to these source to focus on them, while perceptually blending the other local sources out. In the center position of the lab, the listener will receive an encompassing overview of the global and local sound sources.

5. CURRENT PROJECTS

The CRAIVE-Lab, with it's high level of flexibility and multiple avenues for content creation and rendering, has a wide array of ongoing projects from a variety of areas. The following section will highlight three data sonifications that focus on the use the auditory display with a correlated visualization.

5.1. Stock Market Data

A complex set of data that is conventionally approached visually is the monitoring and analysis of current and historic stock market data. There are a variety of resources and websites tailored to graphing and visualizing this information. However, due to the multilayered complexity of the stock market, the data also lend themselves well to an acoustic approach.

The market data of the largest 128 publicly traded United States corporations are analyzed by this system. The relevant information used for the sonification includes each companys daily stock price and share volume traded. The sector of the market that each stock belongs to (e.g. financial, energy, etc.) is also included in the data set.

The sonification occurs in the visual coding environment Max/MSP. The daily stock market information is queried from an online database and loaded into dictionary objects before being

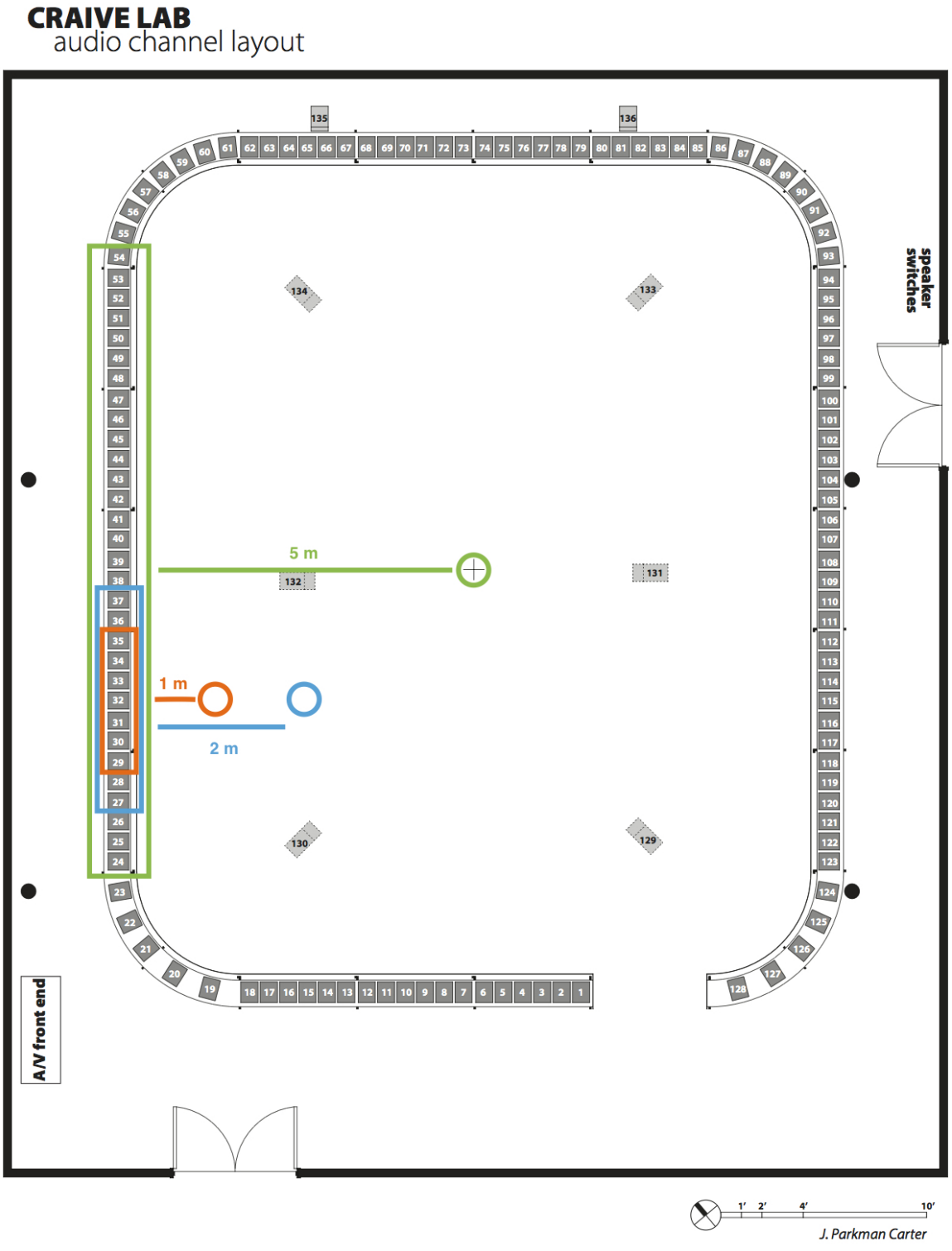


Figure 6: Graphic showing the layout and dimensions of the CRAIVE-Lab, speaker locations, and corresponding channel numbers. Also noted are measurement locations and color-coded channels utilized for each pass [10].

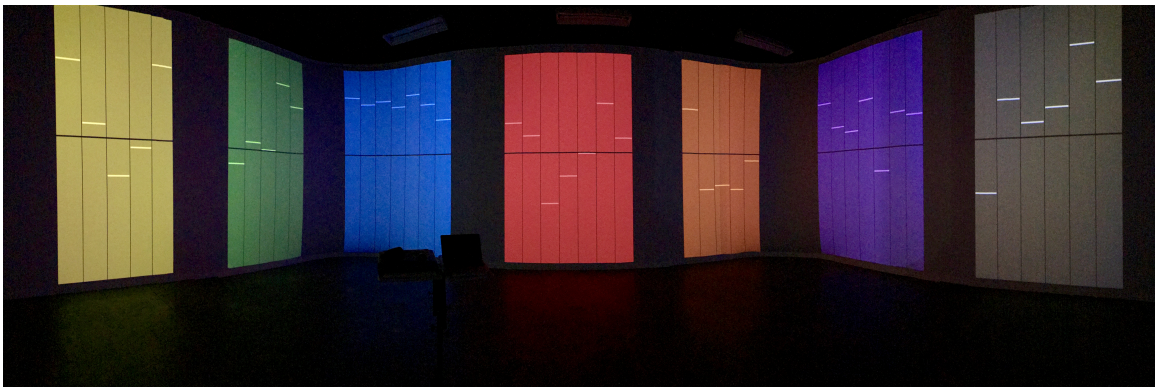


Figure 7: Visualization of stock market trends, where the top 128 publically-traded companies are clustered by sector of the market (e. g. blue-financials, green-energy, purple-technology).

processed. For each stock, the system first determines whether the price experienced a positive or negative percent change over the course of the day. This is then paired with the share volume that was traded. The share volume is mapped to the tempo of a click-train, where a faster tempo indicates a higher traded volume and, conversely, a slower tempo a lower traded volume. Daily volume of these 128 companies is frequently traded on the order of millions to tens of millions of shares. The volume is logarithmically scaled from a volume range of 0 to 50 million to a tempo range of 40 to 200 beats per minute, giving a higher level of granularity in the typical range of trade volume.

The composition of the click-train is determined by the percentage change in each stocks daily price. A negative percentage correlates to a pulse of white noise bursts, while a positive percentage triggers a pulse of sine tones. The pitch of the pulsing sine tone is determined by the percentage change, where a greater gain in the stocks price maps to a higher pitch sine tone. A range of 0 to 2.5 percent is scaled to a midi note range of 60 to 96, or approximately 261 to 2093 Hz. This process is done for all 128 stocks to create a soundscape for judging the quality and performance of the largest companies in the stock market.

Each of the generated auditory streams of the 128 corporations is assigned its own channel in the 128-loudspeaker array. The streams can be sorted into various configurations across the loudspeaker array. The configurations include ascending/descending order by total revenue as well the ability to sort by sector of the market. The system also allows for the isolation of specific sector. Figure 7 shows the generated visualization that accompanies the sonification, which contains companies clustered by their respective sectors of the market (e.g. blue denotes “financials,” green denotes “energy”).

Any available days of historic stock market data are usable. To isolate an interesting period of time in the market, the week of the past 2016 election, November 7 through 11, is loaded in the system. This allows the user to step through each day of the week to listen for and observe clear trends that may (and do) emerge. Clearer patterns are discernible when the corporations are sorted and clustered by sector. For example, corporations in the energy sector may be grouped in loudspeaker channels 13 through 24, while those falling under financials may be grouped about 87 through 98. This creates a distinct separation in the spatial environment, and users can then localize trends that occur amidst a sector while comparing those trends between sectors about the environment.

5.2. Weather Data

Another complex data set that is ever-growing is that of meteorological data. This information is also most often approached from a visual perspective. However, as with stock market data, the multilayered nature of weather data lends itself to a sonification component. The various layers of weather data include temperature, precipitation, wind speed and direction, cloud-level, humidity, UV index. All of these also correspond to current, past, and forecasted conditions [11].

This project also utilizes the CRAIVE-Lab to present an auditory display coupled with a contextual visualization. The data itself is retrieved from the National Climate Data Center of the U.S. Department of Commerce, National Oceanic & Atmospheric Administration (NOAA). Information regarding temperature, precipitation, and wind direction and speed collected at the Albany International Airport creates the foundation for the sonification.

The current weather conditions are used to create the sonification, which is done within Max/MSP. To replicate the current weather conditions, synthetic wind and rain is created. The wind is produced using a pink noise generator multiplied by a sine tone and sent through a resonant filter. The temperature scales the relative fundamental frequency of the wind by changing the sine tone frequency, where a higher pitch denoted a drop in temperature. The rain is created using bursts of white noise filtered to create parabolic pulses. Rumbles resembling thunder are also created with white noise run through a low-pass filter. To create a sense of realism in the audio, recordings at the location for ambient sound were also done. A spatialization of the audio is done by considering the wind direction information. Using cardinal directions mapped over the workspace, the audio stream is concentrated in corresponding loudspeakers. Users are quickly able to discern the information encoded within the sound-the temperature, wind speed and direction, and precipitation levels. This is coupled with a visualization for a complete perceptual experience.

The visualization consists of a time-lapse video taken of downtown Troy, New York in which a large unobstructed view of the sky is shown. The time-lapse was taken over the course of nearly 80 hours, with an image captured every 15 seconds. This totaled 19,142 images. A 48 hour span which saw the greatest di-

versity of weather was exported as a 4 minute video. A small graphical visualization is overlaid on the time lapse video. This compass-like graphic is inspired by the "Wind Wheel of the program Climate Consultant. Wind direction is shown highlighted in 10°increments. The color of the highlight corresponds to the relative current temperature, where dark blue denotes the coldest temperatures recorded through red corresponding to the hottest temperatures.

5.3. EEG Data

Currently being developed is a sonification of EEG data. This data is acquired from patients who experience epileptic seizures. There is believed to be a period of time in the minutes before the onset of a seizure in which observable patterns occur in the brain that could provide a patient with a warning of the incoming episode. An ability for providing such a warning does not currently exist because these observable patterns have not been completely identified. In assisting in recognizing these patterns, an acoustic measure is being attempted.

The EEG data is supplied in 22 separate channels, which correspond to 22 tested locations within the brain. A basic sonification abstraction charts these 22 locations to individual channels about the CRAIVE-Lab. The data is observed in low-voltage readings, which need to be mapped to an audible correlation. A parameter-mapping technique is being explored as an avenue for this data sonification. Each probe location is filtered into the appropriate brain wave frequency bands (alpha, beta, delta, theta) for a more granular analysis. The brain waves receive different timbral qualities in order to distinguish between the four. It is widely believed that indications of an incoming seizure correspond to sudden high levels of correlation between locations in the brain. The use of rhythm is being attempted for conveying these correlations by having pulses for correlated channels temporally align.

6. CONCLUSION

Large immersive environments such as Rensselaer's CRAIVE-Lab have the potential to alter the way people interact with each other and complex sets of data. The ability to explore and collaborate amidst a shared virtual environment at such a scale is a budding and novel approach to an exciting field of research. Providing an equal emphasis on the audio and visual components of the lab will continue to be crucial to develop congruency within a space. Environments that promote collaboration and shared physical presence will continue to enhance human interaction and comprehension of large sets of data.

7. ACKNOWLEDGMENT

This work is supported by the National Science Foundation (NSF #1229391) and the Cognitive and Immersive Systems Laboratory (CISL) at Rensselaer directed Hui Su.

8. REFERENCES

[1] G. Kramer, B. N. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner, and J. Neuhoff, "Sonification report: Status of the field and research agenda," Tech. Rep.

[2] A. S. Bregman, *Auditory scene analysis: the perceptual organization of sound*. Cambridge, Mass.: MIT Press, 1994.

[3] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*, Berlin, 2011.

[4] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[5] J. J. Lopez, M. Cobos, B. Pueo, and C. D. Vera, "Wave-field synthesis : State of the art and future applications," in *Proceedings of the 17th International Conference on Auditory Display (ICAD2011)*, 2011.

[6] T. Höllerer, J. Kuchera-Morin, and X. Amatriain, "The allosphere: A large-scale immersive surround-view in- strument," *IEEE Multimedia*, vol. 16, no. 2, pp. 64–75, 2009.

[7] N. Peters, T. Matthews, J. Braasch, and S. Mcadams, "Spatial Sound Rendering in MAX/MSP with ViMiC," *Proceedings of the International Computer Music Conference*, 2008.

[8] M. Wright, A. Freed, and A. Momeni, "OpenSound Control: State of the Art 2003," *Proc. NIME 2013*, pp. 153–159, 2003. [Online]. Available: papers3://publication/uuid/CEB1F875-E07B-450C-808B-0DC309595D3B

[9] Sennheiser, "Neumann Dummy Head." [Online]. Available: https://www.neumann.com/?lang=en{\&}id=current{_}microphones{\&}cid=ku100{_}description

[10] J. P. Carter, "Craive lab audio channel layout," 2015.

[11] L. Cunningham, "Sonfication and visualization of weather data," 2013.

EXTENDED ABSTRACTS
(POSTER PRESENTATIONS)

BSONIQ: A 3-D EEG SOUND INSTALLATION

Marlene Mathew

New York University,
Music Technology
New York, NY USA
mm5351@nyu.edu

Mert Cetinkaya

New York University,
Music Technology
New York, NY USA
mc5993@nyu.edu

Agnieszka Roginska

New York University,
Music Technology
New York, NY USA
roginska@nyu.edu

ABSTRACT

Brain Computer Interface (BCI) methods have received a lot of attention in the past several decades, owing to the exciting possibility of computer-aided communication with the outside world. Most BCIs allow users to control an external entity such as games, prosthetics, musical output etc. or are used for offline medical diagnosis processing. Most BCIs that provide neurofeedback, usually categorize the brainwaves into mental states for the user to interact with. Raw brainwave interaction by the user is not usually a feature that is readily available for a lot of popular BCIs. If there is, the user has to pay for or go through an additional process for raw brain wave data access and interaction.

BSoniq is a multi-channel interactive neurofeedback installation which, allows for real-time sonification and visualization of electroencephalogram (EEG) data. This EEG data provides multivariate information about human brain activity. Here, a multivariate event-based sonification is proposed using 3D spatial location to provide cues about these particular events. With BSoniq, users can listen to the various sounds (raw brain waves) emitted from their brain or parts of their brain and perceive their own brainwave activities in a 3D spatialized surrounding giving them a sense that they are inside their own heads.

1. INTRODUCTION

Sonification is the method of rendering sound in response to data and interactions and sets a clear focus on the use of sound to convey information [1]. Electroencephalogram (EEG) is the recording of electrical potential from the human scalp containing multivariate data. EEG sonification has been very useful in areas spanning data analysis and medical diagnosis to general-purpose user interfaces in car navigation systems [2]. EEG sonification can give researchers or medical professionals a better idea as to what is happening at a certain location in the brain, when visual analysis can no longer be applied. For example, with fMRI, visual images (scans) of the brain are taken every millisecond, where the analysis of the brain activity takes place after the scan. EEG sonification provides information about brain activity in real-time by providing auditory images that can more easily be interpreted due to their spatial differences [7]. This is one of the main advantages of auditory displays over visual displays. Listening is used as a means to perceive data. Audio feedback for positional control could be very useful in for example, the medical field [5]. Sound is a temporal indicator of the ongoing physical processes in the world around us [16].

This paper presents "BSoniq", a 3-D EEG sound installation, in which, the user can perceive spatial characteristics of EEG signals in a multi-channel environment. With this installation, the users (listeners) wear a wireless EEG headset and listen to sounds generated in real-time from their brain waves to perceive brain activities which they may not be aware of in their daily life. To accomplish a brain electrical activity sonification, brainwave source localization features of multi-channel EEG are converted into sound images. These allow for simple interpretation, because of their spatial temporal differences. Signals recorded from the scalp are "decoded" from the multi-channel EEG, by applying filters and modulation to the EEG signal with an audio file. The main goal is to use sound to render the original data in a suitably transformed way so that we can invoke our natural pattern recognition capabilities to search for regularities and structures. Brainwave sonification is also very practical in brain-computer interface (BCI) user feedback design. Deciding how the control of parameters, processing and filtering of inaudible data are used is important in this process. Using listening as a tool serves both as an aesthetic and/or scientific purpose. The human hearing system is able to decode and interpret complex auditory scenes. The more structured the representation of the sonified data, the better the accessibility and intelligibility of the chosen process [9].

We propose to employ auditory feedback, and thus provide visualization of the brainwaves in the form of spatial sound images, that is, to perform sonification of brain electrical activity. The 14 channels used in this project represent the 14 sensors of the EEG device used.

2. BACKGROUND

Efficient perceptualization of biofeedback or medical data requires a multidisciplinary approach, including the fields of computer science, engineering, psychology and neurophysiology [5]. EEG provides a diagnostically important stream of multivariate data of the activity of the human brain. One of the first attempts of auditory EEG exploration was reported in 1934 by E. Adrian and B. Matthews [15]. They measured the brain activity from a human subject from

https://doi.org/10.21785/icad2017.001

electrodes that were applied to the head, and the channels connected to these electrodes were viewed optically on bromide paper while being directly transduced into sound.
T. Hermann et al. have presented different strategies of sonification for human EEG [3]. Baier et al. used multivariate sonification that displayed salient rhythms as well as used pitch and spatial location to provide cues [15]. Hunt and Hermann conducted experiments to explore interactive sonifications, which they describe as the discipline of data exploration by interactively manipulating the data's transformation into sound [16]. They also realized that the individuality of interacting with sound is important, meaning that one must be able to detect a particular signal even if there are other interfering signals and/or a noisy background present.

There are many experiments converting multi-channel EEG to sound. However, not many use 3D sound to provide spatial cues. Hori and Rutkowski developed an EEG installation, sonifying 14 EEG signals using 5 channels, where the loudspeakers were geometrically located surrounding the listener and termed "A" to "E" from the left to the right [2] on the azimuth angle. By using only five channels, multiple EEG data were combined into one, which was processed and sent to a loud-speaker. This does not allow for details of a specific sensor to be perceived. BSoniq sonifies all 14 channels to speakers located at the azimuth and elevation angles related to an EEG sensor's location for monitoring purposes.

The main areas of EEG sonification are: EEG monitoring, EEG Diagnostics, Neurofeedback, Brain Computer Interface(BCI) feedback and communication as well as EEG mapping to music [11]. BSoniq's main focus is on monitoring or listening. Monitoring generally requires the listener to attend to a sonification over a course of time, to detect events, and identify the meaning of the event in the context of the system's operation [13].

3. HARDWARE

The Emotiv EEG wireless device is used for signal acquisition in this installation. This device has 14 sensors based on the International 10-20 system located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4 (Fig. 1) [14]. The International 10–20 system is an internationally recognized method used to describe the location of scalp electrodes and the underlying cerebral cortex [12]. It was created to ensure a standard format so that the studies of a subject's EEG could be compared over time and subjects could be compared to each other. The "10" and "20" refer to the actual distances between adjacent electrodes that are either 10% or 20% of the total front–back or right–left distance of the skull.

The transmitted wireless EEG signals are received by the Emotiv USB receiver which is connected to a USB port of a PC and sent to Max/MSP for transformation. The data stream coming from the Emotiv device is encrypted by proprietary software, which is subsequently decrypted by Emotiv's SDK. The data is transmitted via the Emotiv's API as raw EEG values in microvolts. The EEG data is then stored as floating point values which are converted from the unsigned 14-bit output from the headset [10].

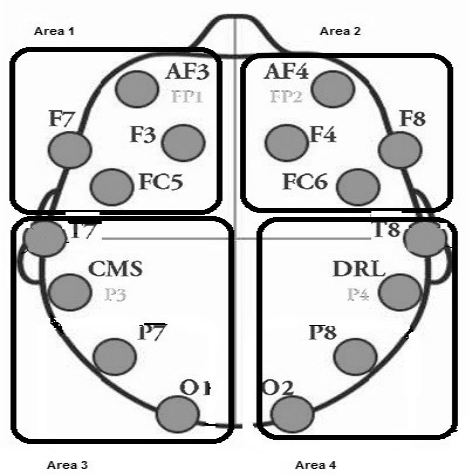


Figure 1: Area division of sensors

Once the EEG signals have been transformed, the sonified data are converted to analog audio signals using audio interfaces and sent to 14 speakers which are geometrically located around the listener. The layout of the speakers represents the layout of the sensors on the user's head, giving the impression that the user is inside his/her own head listening to the various brainwaves in action. The ring topology of the speakers is to provide cues of the azimuth and elevation in the horizontal plane. This, to focus the listener's attention to the correct angle of the sonified signal. Locations of the 14 loudspeakers used in this project are shown in Figure 4. A full sphere setup was used with ten loudspeakers positioned horizontally around the listener and the rest of the speakers were elevated approximately 40 degrees above the listener's head. Details of the sonification process is discussed in the following section.

4. SOFTWARE

Max/MSP, a visual programming language is used for EEG sonification. The actual EEG data transmission is based on Open Sound Control (OSC) protocol, provided by Mind Your OSCs. This is an open source software that sends the raw EEG values received from the Emotiv EEG SDK via User Datagram Protocol (UDP) to Max/MSP for transformation. The sonification is carried out sensor-wise with a sub-patch that receives a single channel EEG signal, which is band-limited and scaled to modulate a sample file. After the modulation takes place, the audio signal is sent via a single channel out to the loudspeaker. For example, the EEG signal of the AF3 channel after being transformed is sent to speaker 10. The full assignments of the EEG channels to the speakers are shown in Table 1. The sonified EEG signal of each electrode is sent to the speaker representing the general location of that electrode on the scalp.

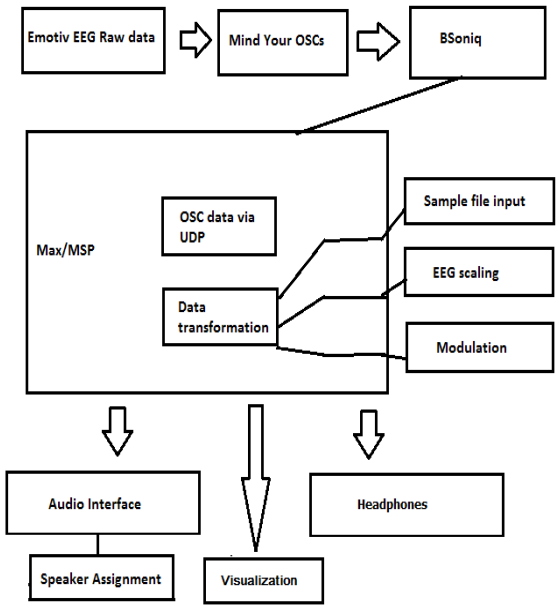


Figure 2: BSoniq flowchart

4.1 Data Representation

When dealing with sonification it is important to choose a specific sound dimension to represent a given data dimension[13]. EEG signals can range from 0.5 to 40 Hz, which makes pitch a good sound attribute to represent any changes in the EEG data. Here, frequency modulation is used to transform the frequency of the EEG signal.

The EEG signal modulation used here is similar to that of a regular Frequency Modulation (FM) synthesis, where you have a carrier signal and a modulating signal. In this project the modulator is the EEG signal and the carrier is the looping sample file.

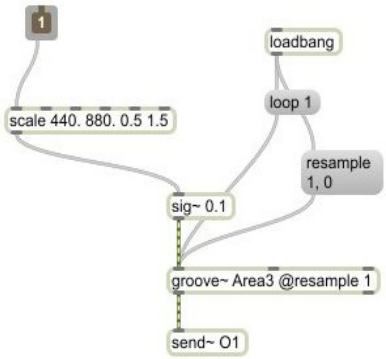


Figure 3: 'test' modulation patch

Table 1: Speaker Assignment

Sensor	Speaker
AF4	1
F8	2
T8	3
P8	4
O2	5
O1	6
P7	7
T7	8
F7	9
AF3	10
F3	11
F4	12
FC5	13
FC6	14

4.2 Scaling

Scaling is important to determine how much the pitch of a sound is used to convey a given change [13]. It shows the relationship between the system and the EEG data. Because each EEG signal may have different frequency and dynamic characteristics, the ability for the user to manipulate the scaling data is important for a better representation of the EEG data and its characteristics.

When the EEG signal is received by the Max/MSP program, it is first limited to 0-4500 uVolts and then scaled from 0-0.5. The scaled EEG signal is then sent to another subpatch 'test' (Fig. 3) to modulate it with a sample file. The sample file (carrier) is a looping audio file excerpt selected by the user. BSoniq provides the user with the flexibility as to what sounds (sample files) to choose for the sonification process, therefor enhancing the listening experience.

Since EEG signals typically range 0.5- 40 Hz, the data here has been scaled by default between 440 – 880 Hz following an octave music model. The user has the option to adjust the range. The data is scaled to values between 0.5 and 1.5. These values determine the amplitude of the modulating signal. The default scaling translates higher EEG values into higher amplitude vectors, and lower EEG values into lower amplitude vectors, which has an effect on the sound output.

4.3 Modulation

At the end of dynamic scaling processes, the modulation process is applied. In BSoniq, the sensors are divided into four areas as shown in Figure 1. The user can choose a different sample file for each area or the same sample file for all four areas. The option for the user to select audio files, allows for a better distinction between the various brain activity levels at sensor level. After this process, the transformed EEG signal is sent to the corresponding loudspeaker. For example, in Fig. 3 the O1 sensor in Area 3 is sonified and sent to loudspeaker 6.

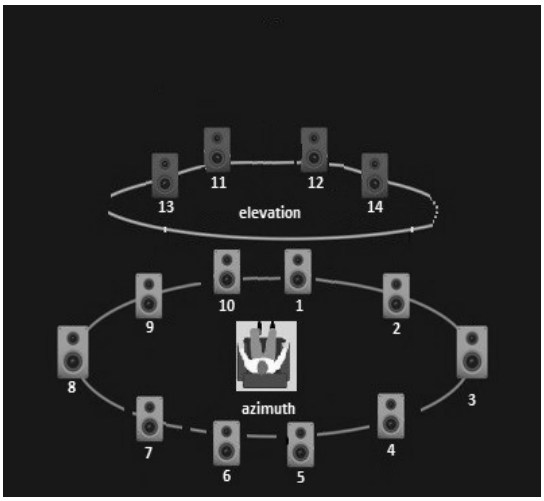


Figure 4: Speaker Layout

4.4 Visual Representation

Though the main feature of BSoniq is its auditory display, it can also visually display the relationship between the activity level of each EEG channel. This optional feature aids the user in visualizing what sensors are active and how much. The visualization is represented by a 3D model head as shown in Fig. 5 and 14 balls representing the EEG sensors and location. These balls, created in Max/MSP/Jitter uses the 'jit.gl.gridshape' object, which generates simple geometric shapes as a connected grid, in this case a sphere. These spherical shapes (balls) increase and decrease in size based on the sensor activity levels. The values used for visualization are the same scaled values used for sonification. The stronger the EEG signal the larger balls become. BSoniq also provides the user with several angles to turn the head model for a better view of the sensors. For example, if the user wants to have a better view of the back sensors, he/she can rotate the head for a side or back view of the balls (sensors).

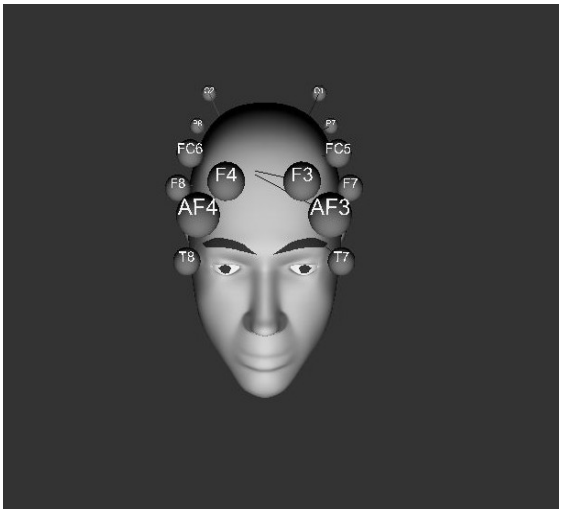


Figure 5: Visual representation of sensor activity

4.5 Headphones

As described in the previous sections, BSoniq was first designed as an installation for a 14-channel loudspeaker setup. However, most people do not have a 14-channel loudspeaker system accessible to them. Since this installation was intended for general use and the ability to convolve each modulated EEG signal with a corresponding loudspeaker impulse response (IR), brought on the idea of taking this project to the next level. Users would be able to experience BSoniq with a pair of headphones allowing flexibility for its use in non-laboratory environments.

In the binaural format of BSoniq, the approach is to start with measuring impulse responses of each loudspeakers that are geometrically setup to represent the 14 sensor locations of the Emotiv EEG device using the Neumann KU-100 Dummy head. The resulting recorded stereo impulse responses are then split into left and right channels using Matlab, yielding a total of 28 impulse responses. Max/MSP is used here to provide the binaural experience in real-time.

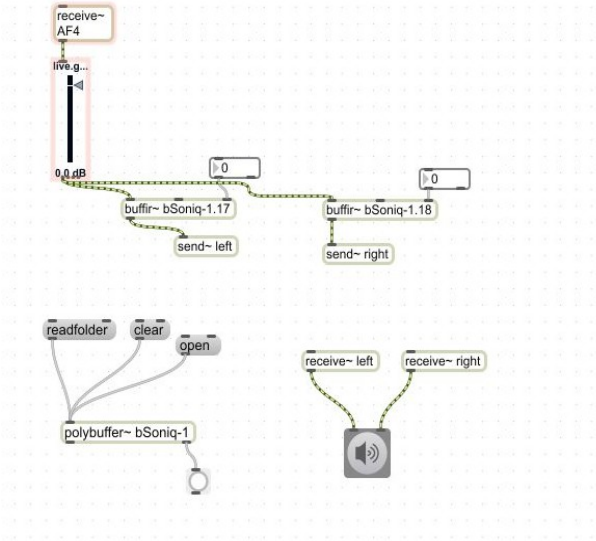


Figure 6: Partial HRTF patch

In Max/MSP, we used the “buffir~” object, which is described as “a buffer-based FIR filter that convolves an input signal with samples from an input buffer” [16]. In this case the EEG signal is convolved with the corresponding IR. Since we are dealing with many channels of audio, we used the “polybuffer~” object to ease the process of loading and delivering the IRs to corresponding “buffir~” objects. After each EEG convolution, the output is sent to the corresponding left or right channel, thus giving a virtual representation of the various EEG sensor locations. To give an example, in Figure 6, the signal from the AF4 sensor is sent to the buffir~ object, which takes the left output signal and convolves it with IR sample 17 and convolving right output signal with IR sample 18. After the convolution of the AF4 signal, both left and right convolved signals are sent to the “ezdac~” object for output.

5. PROTOTYPING

The prototyping phase included evaluation by five users; three males and two females. Most users were able to distinguish EEG frequency changes as a result of the modulation process. Users did indicate that using a sample file of sound they were very familiar with, helped them to quicker understand the sonification process. For example, one user could not distinguish much difference with a sample file that was a bell sound, but was able to pick up on subtle frequency changes with a sound sample that was a snippet of a song he knew very well. While wearing headphones, users were also able to hear the location from where a particular modulated EEG signal was coming from, indicating the virtual placement of the sensors.

6. DISCUSSION

We presented BSoniq, which uses multi-channels to sonically represent EEG data in real-time 3D space using frequency modulation. BSoniq could be used for both online and offline sonification. By applying filters and parameter controls, it is possible for the user to focus on the area of interest within the signal. This is useful for real-time applications like EEG monitoring or EEG feedback. The inclusion of 360 degrees spatial cues permits the parallel sonification of many or all electrodes without losing clarity in the display. Clarity of the sonification, however also depends on the strength of the EEG signal capture from the device. The signal could also contain artifacts, which could be reduced or removed in order to yield a clearer signal for the display. What also needs to be noted is that the perceptual capabilities of the listener is important. If the listener is unable to distinguish sounds or incapable of hearing certain frequencies, then this would affect the user's perception of the installation's functionality.

Future work includes conducting additional evaluations for necessary design improvements as well as upgrading BSoniq to include other popular EEG devices. The current installation only allows for the user to remain stationary. Allowing the user's head movement and tracking, is a feature that will be added to create a fully integrated system.

To conclude, we believe we accomplished our goal of EEG sonification using 3D spatial cues. Even though BSoniq started out as an installation, mainly for an aesthetic user listening experience, we also believe that in addition to sonification, the visualization component could also be enhanced into an artistic EEG visualization application using geometric data and transformations for artistic applications. That is, exploring methods that uses OpenGL for example, to create 3D spatial-spectral representations of an EEG signal.

7. REFERENCES

[1] Thomas Hermann, Andy Hunt and John Neuhoff. “Auditory Display and Sonification”. *The Sonification Handbook*, 2011.

[2] Gen Hori and Tomasz M. Rutkowski. “Brain listening–a sound installation with EEG sonification”. *Journal of the Japanese Society for Sonic Arts*, 4(3):4–7, 2000.

[3] Thomas Hermann and Helge Ritter. “Listen to your data: Model-based sonification for data analysis”. *Advances in intelligent computing and multimedia systems*, 8:189–194, 1999.

[4] Stephen Barrass and Gregory Kramer. “Using sonification”. *Multimedia systems*, 7(1):23–31, 1999.

[5] Emil Jovanov, Dusan Starcevic, and Vlada Radivojevic. “Perceptualization of biomedical data”. IN MEDICINE, page 189, 2001.

[6] Teruaki Kaniwa, Hiroko Terasawa, Masaki Matsubara, Tomasz M Rutkowski, and Shoji Makino. “EEG auditory steady-state synchrony patterns sonification”. In *Signal & Information Processing Association Annual Summit and Conference* (APSIPA ASC), 2012 Asia-Pacific, pages 1–6. IEEE, 2012.

[7] Tomasz M Rutkowski. “Multichannel EEG sonification with ambisonics spatial sound environment”. In *Asia-Pacific Signal and Information Processing Association*, 2014 Annual Summit and Conference (APSIPA), pages 1–4. IEEE, 2014.

[8] Tomasz M Rutkowski, Francois Vialatte, Andrzej Cichocki, Danilo P Mandic, and Allan Kardec Barros. “Auditory feedback for brain computer interface management–an EEG data sonification approach”. In *Knowledge-Based Intelligent Information and Engineering Systems*, pages 1232–1239. Springer, 2006.

[9] Timothy Schmele and Imanol Gomez. “Exploring 3d audio for brain sonification”. In *International Conference of Auditory Display*, 2012.

[10] Haracio Tome-Marques and Bruce Pennycook. “From the unseen to the s[cr]een eshofuni, an approach towards real-time representation of brain data”. 2014.

[11] A Väljamäe, T Steffert, S Holland, X Marimon, R Benitez, S Mealla, A Oliveira, and S Jordà. “A review of real-time EEG sonification research”. In *International Conference of Auditory Display*, 2013.

[12] Neuroscience For Kids. (n.d.)., from <http://faculty.washington.edu/chudler/1020.html> Retrieved February 10, 2016.

[13] Walker, Bruce, and Nees, Michael. "Theory of sonification."*The Sonification Handbook*: 9-39, 2011.

[14] Emotiv Epoc EEG. <https://www.emotiv.com>. Retrieved July 13, 2015.

[15] Gerold Baier, Thomas Hermann, and Ulrich Stephani. “Multi-channel sonification of human EEG”. I*n Proceedings of the 13th International Conference on Auditory Display*, 2007.

[16] Max/MSP/Jitter Graphic software development environment. Cycling '74. <www.cycling74.com>. Retrieved November 3, 2016.

MULTISENSORY CUE CONGRUENCY IN THE LANE CHANGE TEST

Yuanjing Sun[1], Jaclyn Barnes[2], Myounghoon Jeon[1,2]

Mind Music Machine Lab
Michigan Technological University,
[1]Department of Cognitive and Learning Sciences, [2]Department of Computer Science,
1400 Townsend Dr. Houghton, MI 49931 USA
{ysun4, jaclynb, mjeon}@mtu.edu

ABSTRACT

Drivers interact with a number of systems while driving. Taking advantage of multiple modalities can reduce the cognitive effort of information processing and facilitate multitasking. The present study aims to investigate how and when auditory cues improve driver responses to a visual target. We manipulated three dimensions (spatial, semantic, and temporal) of verbal and nonverbal cues to interact with visual spatial instructions. Multimodal displays were compared with unimodal (visual-only) displays to see whether they would facilitate or degrade a vehicle control task. Twenty-six drivers participated in the Auditory-Spatial Stroop experiment [1] using a lane change test (LCT). The preceding auditory cues improved response time over the visual-only condition. When dimensions conflicted, spatial (location) congruency had a stronger impact than semantic (meaning) congruency. The effects on accuracy was minimal, but there was a trend of speed-accuracy trade-offs. Results are discussed along with theoretical issues and future works.

1. INTRODUCTION

For decades, in-vehicle technologies have rapidly increased. Given that vision is fully occupied while driving, technologies using other modalities, such as speech recognition or vibrotactile notifications, have become pervasive in vehicles. Multiple Resource Theory (MRT) [2] indeed supports the use of multimodal interfaces in the vehicle context. However, despite potentially allowing drivers to process more information in parallel, multimodal interfaces still occupy attentional resources. Does more information always mean more facilitation? With bad design, multimodal displays might cause information overload or degrade performance, which could lead to safety hazards on the road. For example, suppose that the personal navigation device (PND) tells a driver to make a left turn, but at the same time, the collision warning system alerts the driver that there is a hazard coming from the left lane. How would the driver respond to this conflicting information? Even though multimodal displays might benefit a single task, they might not always benefit multiple tasks, especially when modalities conflict with one another at the same time. The present study aims to address this issue to identify underlying mechanisms and provide design guidelines for in-vehicle multimodal-visual and auditory-displays.

1.1. Multiple Resource Theory

Wickens' MRT [3] has been used to predict or analyze interference between concurrently perceived signals. Two tasks that demand *separate* levels (e.g., one visual and one auditory tasks) will interfere with each other *less* than two tasks that both demand one level of a given dimension (e.g., visual and visual tasks). It provides a basic theoretical endorsement to the blooming implementation of multimodal interfaces. However, MRT is also challenged by multisensory illusions, such as McGurk illusion or Ventriloquist Illusion [4], where information from different channels are synthesized into a new, distinct signal. The conflict between MRT and multisensory illusion prompts a more detailed examination of how multisensory perception influences information processing.

1.2. Multimodal Benefits

Multimodality provides synergy at the cost of significantly less cognitive effort than processing information from a single modal channel [5]. By providing processing advantages for grouping and organizing signals with the lowest workload, redundancy in multimodal display can increase the bandwidth of concurrent information processing. Here, the arrangement of multimodal signals becomes decisive to the occurrence and strength of multimodal benefits. The degree of multimodal benefits follows both (1) spatial rules and (2) temporal rules. However, several conflicting studies make it difficult to identify exactly from where the facilitation derives.

1.2.1. Spatial rules

In his review of crossmodal spatial attention [6], Spence proposed the performance benefit on ipsilateral (on the same side) cued trials over contralateral (on the opposite side) cued trials. A possible mechanism for this might be "spatial proximity" between stimulus and response. In other words, a spatially predictive auditory or visual cue would always lead to an exogenous attentional shift and narrow down spatial attention to the cue direction. A spatially corresponding mapping of left stimuli to left responses and right stimuli to right responses yielded better performance (i.e., faster reactions and fewer errors) than a spatially incongruent mapping [7].

https://doi.org/10.21785/icad2017.015

1.2.2. Temporal rules

There are divergent research outcomes on the temporal interval between auditory cue and visual target. For example, crossmodal synesthesia [8] predicts a synchrony benefit. It claims that the responses to multimodal cues will benefit when there is a maximum overlap between cue and target. In contrast, Posner's spatial cuing task proposes a "preparation function", suggesting that the response time would become fastest when a priming tone was 200 ms ahead of the visual target [9]. In this line, the present study selected 200 ms as preceding timing as the asynchrony condition to contrast with the synchrony condition.

1.3. Type and Demand of Visual Tasks

Multimodality does not always provide benefits over unimodality. Sinnett, Soto-Faraco and Spence [10] manipulated perceptual load (frequency of visual targets) and working memory load (numbers of responses) to compare the redundant gain of multimodality. The result indicated that both multisensory facilitation and inhibition can be demonstrated by changing the task type and visual demand.

In particular, Wickens and other researchers [11] suggested that a redundant auditory display may facilitate a visual scanning task but not an ongoing visual tracking task. In audiovisual redundancy studies, ongoing visual tracking tasks require continuous visual attention. In the context of visual tracking tasks, there are periodic interrupting tasks that are discrete in nature. A meta-analysis of 29 studies [12] comparing visual-auditory tasks with visual-visual tasks has shown that auditory presentation for a discrete task resulted in a significant 15% performance advantage over visual-only presentation. In particular, the auditory advantage increased when the two visual inputs were end-to-end. In other words, the auditory cues were more helpful when the interval between two visual inputs was shorter (i.e., visual perceptual load is high). It can also be inferred that the auditory-visual facilitation would occur in visually-demanding tasks (e.g., the demand of the visual scanning task is higher than the visual tracking task). The lane change test includes both visual scanning (identifying a visual target) and visual tracking tasks (maintaining lane position). We anticipate the use of auditory cues will be more helpful for the visual scanning task than the visual tracking task.

1.4. Auditory-Spatial Stroop Task

Inheriting from the original color-word naming Stroop paradigm, researchers utilized the Auditory-Spatial Stroop task to investigate location-meaning conflicts in multimodal processing. Auditory-Spatial Stroop task, originally introduced by Pieters [13], consists of directional verbal cues presented congruently or incongruently with a visual target. Mayer and Kosson showed that there was a significant lag in reaction time (RT) to the target location when incongruent auditory cues were presented. However, incongruent visual cues did not delay RT to auditory targets [14]. It suggested that a visual distractor is easier to ignore than an auditory distractor. The asymmetric anti-distraction feature between vision and audition indicates that the modality of message in multitask signaling could interfere with the priority level in response selection. Barrow and Baldwin [15], [16] used the Auditory-Spatial Stroop task to simulate the potential location-meaning conflict that might happen under several multimodal in-vehicle devices (e.g., side collision avoidance warning and PND). For example, the word, "left" or "right" is presented in a congruent or incongruent position with its meaning. They found that it is more difficult to ignore the spatial information of the verbal cues than the semantic information when there was a location-meaning conflict.

2. THE CURRENT STUDY AND HYPOTHESES

Understanding different mechanisms involved in multisensory perception is important for choosing appropriate modalities to convey messages for certain tasks. Designers need to have an overall consideration of the implementation environment and priority schedule of all the tasks. The present study intends to ascertain the decisive mechanism(s) in multisensory perception. Since the interference in spatial, semantic, and temporal dimensions is not always orthogonal, the interference of the three dimensions was respectively compared with the visual-only condition. In the view of this research purpose, we constructed three major sets of hypotheses:

- Hypotheses 1: Spatial Rules
  - H1a: Spatially congruent audio-visual (A-V) pairs will have shorter RT than the visual-only condition.
  - H1b: Spatially incongruent A-V pairs will have longer RT than the visual-only condition.
  - H1c: If two above are true, it could be inferred that spatially congruent A-V pairs will have shorter RT than spatially incongruent A-V pairs.
- Hypotheses 2: Temporal Rules
  - H2a: Asynchronous (i.e., preceding auditory cues) A-V pairs will have shorter RT than the visual-only condition.
  - H2b: Synchronous A-V pairs will not have longer RT than the visual-only condition.
- Hypotheses 3: Spatial-Semantic Conflict
  - H3a: Spatiality will have a stronger impact than semanticity. Spatially incongruent and semantically congruent conditions will have longer RT than the visual-only condition.
  - H3b: Spatially congruent and semantically incongruent conditions will have shorter RT than the visual-only condition.

3. METHOD

3.1. Participants

Twenty-six participants (23 male, 3 female; M_{age} = 20.6, SD_{age} = 2.3; $M_{YearsOfDriving}$ = 4.5, $SD_{YearsOfDriving}$ = 2.86) were recruited from the undergraduate participant pool of an American technical university. Participants were native English speakers at least 18 years old. To control for driving skill, participants were required to possess a valid driver's license and have at least 2 years of driving experience.

	Nonverbal Cue		*Verbal Cue*			
Spatial	*Congruent*	*Incongruent*	*Congruent*		*Incongruent*	
Semantic	*N/A*	*N/A*	*Congruent*	*Incongruent*	*Congruent*	*Incongruent*
Synchronous	Track 1	Track 2	Track 3	Track 4	Track 5	Track 11
Asynchronous	Track 6	Track 7	Track 8	Track 9	Track 10	Track 12

Table 1: Summary of audio-visual stimulus mappings in spatial, semantic, and temporal dimensions for various tracks used in the experiment. Tracks had 78% intended cues and 22% distractor cues to prevent participants from anticipating actions based on the audio cues.

3.2. Stimuli

3.2.1. Visual Stimuli

Each track began with a START sign, then 18 sets of lane change signs, and ended with a FINISH sign. The lane change signs appeared in an overhead position on a gate or bridge over the simulated roadway. They were composed of one down arrow and two Xs in three separate black boards (shown in Figure 1).

Figure 1: Visual target stimulus used in OpenDS lane change test scenarios.

3.2.2. Auditory Stimuli

Two non-verbal stimuli and four verbal stimuli were used as auditory cues in twelve tracks out of fourteen in total. The two non-verbal stimuli were a single and a double beep, indicating a single or double lane change. The non-verbal stimuli had no semantic congruency, but had spatial and temporal congruency.

The four verbal cues were "LEFT", "RIGHT", "LEF-LEFT", and "RIGH-RIGHT". The auditory stimuli were normalized to equal duration of 350ms at 60 dB level. The length and loudness of auditory cues were determined by reference to similar demands of the perceptual-motor experiments conducted in previous research [17]. All auditory stimuli were presented at a level of approximately 60 dB from the JVC-HA/RX300 stereo headset. The speech clips "LEFT" and "RIGHT" were recorded using the free online Text-to-Speech (TTS) service (www.naturalreaders.com) at medium speed with a female voice (Laura, US English).

Sped up verbal clips "LEF-LEFT" and "RIGH-RIGHT" indicated the direction of double lane changes (i.e., from the left most lane of the three-lane simulated road to the right most lane or vice versa). These clips were created by importing the original TTS files "LEFT" to Audacity 2.1.0 and replicated the word "LEFT" to two audio tracks. For the first audio track, the first part "LEF" was kept and for the second audio track, the full word "LEFT" was kept. Finally, the two audio tracks were combined and compressed to 350 msec. The "Change Tempo" effect in Audacity was used to adjust the length of audio clip without changing the pitch.

In addition to temporal congruency, verbal cues had spatial congruency and semantic congruency. Thus, the mapping relationship of verbal cues with visual targets had both spatial congruency (physical location of the verbal cue to visual indication) and semantic congruency (meaning of the verbal cue to visual indication). For example, consider a semantically congruent and spatially incongruent condition. Given the visual cue for a single lane change to the left, the participant hears a verbal cue "LEFT" coming from the right speaker.

3.3. Driving Scenario and Apparatus

The Auditory-Spatial Stroop experiment was developed on the basis of the embedded ReactionTest scenario in OpenDS 2.5. We re-implemented the Lane Change Test Toolkit in OpenDS and made modifications according to ISO26022-2010. The original Lane Change Test (LCT) [18] is a simple laboratory dynamic dual task method, which quantitatively measures performance degradation in a primary driving task. In this way, researchers can manipulate the timing and multimodal combinations of lane change signs to capture different driving patterns under different conditions. The simulator consisted of SimuRide software with a 39" monitor and steering wheel. Speed was fixed and the pedals were not used. The primary task required a participant to drive in a straight three-lane road containing a series of lane changes defined by visual targets (Figure 1). In the original LCT, the simulated track length is 3,000 m, corresponding to three minutes of driving at a constant 60 kph [19]. However, to increase the perceptual workload in the primary driving task, the speed in the current experiment was increased to 110 kph (70 mph), a freeway speed limit in some US states. The 18 lane change signs were distributed at intervals of approximately 150 meters. In other words, each lane change maneuver needed to be completed within roughly 4 seconds. The lane change signs were made visible approximately 40 meters before the sign position. In this way, the lane keeping maneuver distinguished two successive lane change maneuvers and provided a buffer if participants made an erroneous lane change at the previous sign. The deviated

Order	Track Number													
1	0	8	2	11	6	4	7	1	10	13	5	12	3	9
2	0	9	3	12	5	10	1	7	4	13	6	11	2	8
3	0	7	4	6	12	2	8	9	3	13	5	11	10	1
4	0	1	10	11	5	3	9	8	2	13	12	6	4	7

Table 2: Summary of the partially counterbalanced exposure orders for the four participant groups. Visual-only baseline tracks 0 and 13, which remained constant for all groups, are highlighted in grey.

distance from the last sign did not influence the start position of the upcoming sign.

3.4. Experimental Design

Nonverbal tracks had two dimensions: spatial and temporal. Since the visual target appeared in every track, it was the reference for "congruent" or "incongruent". In the spatial dimension, a condition is congruent if the cue is played from the same side the driver should move toward, e.g., a single tone from the left side when the visual cue also indicates a single lane change to the left. In the temporal dimension, the asynchronous condition meant that the auditory cues appeared 200 ms ahead of the visual target. The synchronous condition indicated no temporal gap between audio-visual stimuli.

Verbal cues had three dimensions, spatial and temporal plus an added semantic dimension. Congruency in the semantic dimension refers to a match between the meaning of the word(s) in the cue and the desired maneuver.

The experiment was a 2 (spatial congruency) * 2 (semantic congruency) * 2 (temporal congruency) within-subjects design. All conditions are shown in Table 1. Apart from the twelve conditions, participants were given two chances of the baseline (visual-only) tracks, separately numbered as Track 0 and Track 13. The Track 13 were inserted between the 9th and 11th run to see the trend of the learning effect. Aside from the visual-only tracks, each track included 78% target cues and 22% distractor cues to prevent participants from anticipating maneuvers from the auditory cues. The order of 14 tracks was partly counterbalanced as shown in Table 2. Participants were randomly distributed into four groups. Orders 1 & 2 were reversed sequential orders. Order 3 split the tracks in the middle to the two extremes. Order 4 was the reversed sequence of Order 3. In this way, the order effects were minimized. To reduce participants' adaptation to repeated patterns, asynchrony, congruency, and modality were considered in each order.

3.5. Procedure

After signing a consent form, participants watched an instructional video for an overview of the experiment and guidance on how to maneuver the lane change test. The primary task in the LCT was to rapidly change lanes as directed by the visual targets and to maintain the center of the lane between maneuvers. At the same time, unpredicted auditory cues were sent out via the headset. The participants were required to count all auditory cues based on their locations (either left or right ear) and reported the subtotal number of each side to the experimenter at the end of each track.

Before the experiment started, an equivalent hearing test and training trial were given to the participant to make sure that all cues were recognizable to all participants. Also, the experimenter ensured that all participants comprehended the tasks in the whole process of the experiment. A RT histogram displayed briefly after the completion of each track. As long as the participant reached 50% accuracy, they were considered qualified to enter the formal portion of the experiment.

3.6. Metrics

Reaction time (RT) and percentage of correct lane changes (PCL) were two direct metrics for speed and accuracy [11]. The car position parameters (i.e., positional coordinates) were automatically recorded by the driving simulator at the sampling rate of 10 Hz [19]. The reaction to the stimulus was measured as the time span between stimulus and a steering wheel angle outside of the ordinary range for lane keeping. The reaction timer was activated simultaneously with the earlier cues' appearance and ceased when the car maintained the targeted lane for 800 ms. The 800 ms was then subtracted from the reaction timer's value, leaving the true reaction time as the output. The maximum RT window for correct completion of a lane change was either 4.1 seconds or 117 meters after the lane change sign (OpenDS Reaction Task default settings). Otherwise, it was recorded as an incomplete lane change. The reaction timer also excluded overshooting the target lane from recordings of correct lane change maneuvers.

The accuracy was the percentage of correct lane change (PCL) in each track. The correctness of lane change was defined by the driver's position before and after the lane-change maneuver [20]. For each road segment between two signs, the lane where the vehicle was most frequently positioned was identified. Consistent lane choices were then defined as those cases where the vehicle remained in the lane for more than 75% of the segment. This selected lane was then compared to the correct target lane. For each track, the PCL was then calculated as the fraction of the consistent lane

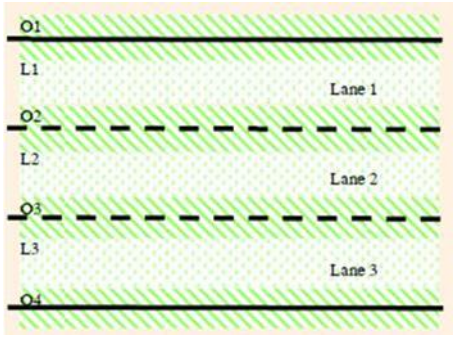


Figure 2: A diagram of effective area for reaction timer to distinguish correct lane change maneuver from erroneous or no lane change in LCT scenario [20].

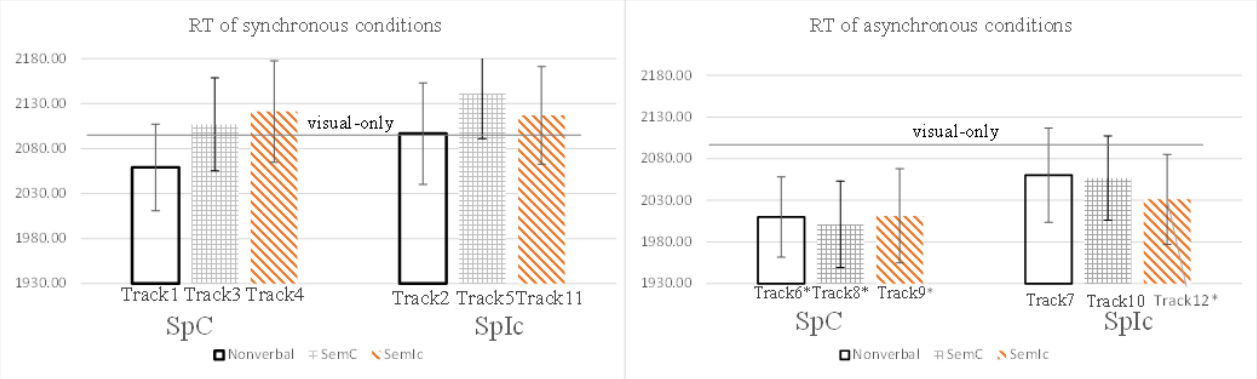


Figure 3: The left half plot is the average RT of synchronous conditions and the right half plot is the average RT of asynchronous conditions. Visual-only is marked as a baseline. Abbreviations used in the graph include synchronous (Syn), asynchronous (Asyn), spatial (Sp), semantic (Sem), congruent (C), and incongruent (Ic).

choices that were correct.

To determine this position, the 3-lane road was divided into different zones, corresponding to parts of the lanes as Figure 2 shows. The dotted zones L1 to L3 corresponded to a correct position in lane 1 (left lane), lane 2 (center lane) or lane 3 (right lane), while the stripe zones "O" corresponded to out of lane positions. The lateral position of the driver was defined by the zone which contained the 75% of his/her trajectory between two signs. If not, then the position was considered as being out of lane and the reaction timer outputted an NA instead of RT. The correctness of each lane change was defined as follows: (1) "Correct LC": the end position of the driver was in the intended lane; (2) "No LC": the driver was in the same Li zone at start and end positions; and (3) "Erroneous LC": the end position of the driver was in a different lane than both the starting lane and the target lane.

4. RESULTS

For planned comparisons, familywise Type I error rate is generally deemed unnecessary [21]. Thus, Bonferroni correction was not applied to the alpha level in the following paired samples t-tests. Twelve paired samples t-tests on RT and accuracy were respectively conducted to examine the mean difference between each condition track and the visual-only condition (mean of the two visual-only tracks).

Figure 3 shows average RT of correct lane-changes across all conditions with standard error bars. The visual-only condition is the baseline to mark facilitation versus deterioration. The unit of y axis is milliseconds. The asterisks in Tracks 6, 8, 9, & 12 show significant differences in paired samples t-tests when compared to the visual-only condition. For tracks with nonverbal cues, the asynchronous spatially congruent condition (Track 6) $t(26) = -2.383$, $p = 0.025$ showed significantly faster RT than the visual-only condition. For tracks with verbal cues, the asynchronous spatially congruent semantically congruent condition (Track 8) $t(25) = -2.478$, $p = 0.02$, the asynchronous spatially congruent semantically incongruent condition (Track 9) $t(25) = -2.817$, $p = 0.009$, and the asynchronous spatially incongruent semantically incongruent condition (Track 12) $t(25) = -2.665$, $p = 0.013$ showed significantly faster RT than the visual-only condition.

Figure 4 shows average accuracy in 12 conditions. For accuracy, there was no clear results or patterns, but synchronous conditions tended to show higher accuracy than asynchronous conditions.

Since the visual-only condition served as the baseline in comparison with all conditions, the subtraction of multimodal tracks over the visual-only tracks are denoted as ΔRT and Δ% in RT and accuracy respectively between multimodal tracks and visual-only tracks. This simplified version of the twenty-four paired samples t-test results is used in the discussion.

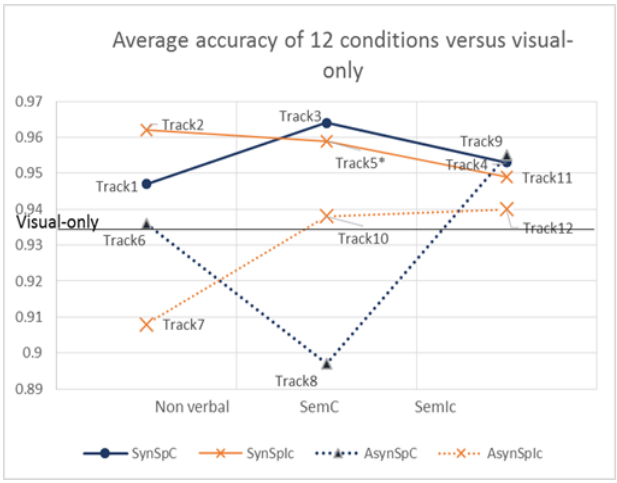


Figure 4: Average accuracy of 12 conditions versus visual-only. Abbreviations used in the graph include synchronous (Syn), asynchronous (Asyn), spatial (Sp), semantic (Sem), congruent (C), and incongruent (Ic).

5. DISCUSSION

The present experiment used the Auditory-Spatial Stroop paradigm [13] in a lane change test scenario to measure the variance of driving performance under the manipulation of spatial, semantic, and temporal congruency of auditory and visual cues.

5.1. Spatial Rules (H1)

The results showed that spatially congruent conditions, at least in the asynchronous conditions (Tracks 6, 8, & 9), had significantly faster RT than the visual-only condition. This partially supported H1a. It demonstrated that spatially congruent A-V association enhances visuospatial response speed. As with the spatial rules in multimodal facilitation, it

		Nonverbal Cue		*Verbal Cue*			
	Spatial	*Congruent*	*Incongruent*	*Congruent*		*Incongruent*	
	Semantic	*N/A*	*N/A*	*Congruent*	*Incongruent*	*Congruent*	*Incongruent*
Synchronous	*ΔRT*	-32.31	5.28	15.51	50.06	29.82	25.46
	Δ%	1.20%	2.70%	2.90%	1.80%	2.50%	1.40%
Asynchronous	*ΔRT*	-81.64*	-31.17	-90.16*	-80.11*	-34.69	-60.27*
	Δ%	0.10%	-2.70%	-3.70%	2.00%	0.30%	0.50%

Table 3: Subtraction of conditional RT and accuracy out of baseline RT and accuracy. Asterisks indicate statistical significance.

is easier to direct attentional focus in different sensory modalities to the same spatial location rather than different location [22]. However, the mixed results in the spatially incongruent conditions (even track 12 shows significantly faster RT than the visual-only) seem to show the several sources of confounding effects on RT. Therefore, the comparison of incongruent multimodal tracks and visual-only tracks did not support H1b that incongruent multimodal cue-target pairs will have longer RTs than those in the visual-only condition. Rather, all asynchronous conditions tended to show faster RT. This might be because sound's arousal effect increased drivers' attention level and thus, sped up the drivers' RT regardless of whether the sounds are related to the primary driving task or not [6]. The arousal effect might somehow cancel out the spatially incongruent cues' plausible delay effects. Overall, the data tend to support H1c as shown in Figure 3.

5.2. Temporal Rules (H2)

H2a and H2b are concerned with the temporal rules in crossmodal links. As hypothesized in H2a, the asynchronous multimodal pairs (Tracks 6, 7, 8, 9, 10 & 12) showed shorter RT than the visual-only baseline, either significantly (Tracks 6, 8, 9, & 12) or numerically (Track, 7 & 10). Therefore, H2a seems to be mostly supported by the results. The results support Posner's preparation function theorem [9] that priming auditory cues benefit reaction time. In H2b, we hypothesized that RT in the synchronous multimodal pairs would not be longer than that in the visual-only condition (based on crossmodal synesthesia). The majority of synchronous pairs (Tracks 2, 3, 4, 5, & 11) showed numerically longer RT than visual-only condition. This trend seems against H2b. Thus, results of RT did not support the synchrony benefit predicted by crossmodal synesthesia.

Why did crossmodal synesthesia not occur in this experiment? The Colavita bias [23] might be the reason. The Colavita visual dominance effect refers to the phenomenon where participants respond more often to the visual component of an audiovisual stimulus, by commonly neglecting the auditory component. In speeded audiovisual asynchrony discrimination tasks, Koppen and Spence investigated the influence of different Stimulus Onset Asynchrony. To many synchronous A-V pairs, the visual cue was actually perceived 12ms faster than the auditory cue which might lead to a prior-entry effect. In summary, generating auditory cues at the same time as visual cues might not result in the simultaneous processing necessary for crossmodal synesthesia.

5.3. Spatial-Semantic Conflict (H3)

In his spatial cuing task, Posner used only the non-verbal sound as auditory cues. The present experiment expanded the asynchrony benefit to the verbal cues. The addition of verbal cues created an interesting case: spatial and semantic conflict. The asynchronous (200 ms in this experiment) A-V pairs sped up response time either when there was no location-meaning conflict between A-V modalities (Tracks 8 & 12) or when the auditory cues were only spatially congruent with the visual target (Tracks 8 & 9). For the tracks having verbal cues, the spatially and semantically congruent groups had the shortest RTs among verbal pairs. (Track 3 had faster RT than Tracks 4, 5, & 11. Track 8 has faster RT than Tracks 9, 10, & 12).

H3a predicted that spatially incongruent and semantically congruent pairs would have longer RT. This was partly supported by Track 5. Track 5 showed the longest RT. Track 10 did not support this hypothesis, perhaps because its asynchrony improved RT. On the other hand, H3b predicted that spatially congruent and semantically incongruent pairs would have shorter RT. This was also partly supported by Track 9, which showed significantly faster RT than the visual-only. Track 4 did not support this hypothesis, perhaps because its synchrony degraded RT. Taken together, spatiality seems to be more powerful than semanticity in both cases (i.e., "where" information is more rapidly processed than "what" information is). However, the temporal dimension seems to have priority and confounds the results.

One interesting result came from spatially incongruent semantically incongruent pairs (Tracks 11 & 12). These had better performance than spatially incongruent and semantically congruent pairs (Tracks 5 & 10) because the spatial and semantic nature within the verbal cues were still consistent with each other despite being incongruent with the visual cue (e.g., visual cue directing the right, but auditory cue saying the word "LEFT" coming from the left speaker). The conflict within the verbal cue seems to have stronger effects than the conflict between A-V modalities.

5.4. Task Type and Speed-Accuracy Tradeoffs

Overall, the effects of the auditory cues on accuracy was very small. This could be explained by the distinction between visual scanning and visual tracking. Identifying the visual indication could be considered a visual scanning task. After changing the lane, keeping the lane position (by definition of PCL) could be the visual tracking task. As expected, auditory cues influenced the visual scanning task (reaction time) more than the visual tracking task (accuracy). However, there was also a trend of typical speed-accuracy tradeoffs. Most

asynchronous auditory conditions improved reaction time, but most asynchronous auditory conditions seem to have lower accuracy than the synchronous auditory conditions. Triggering the response faster does not guarantee better or smoother control of the vehicle. Therefore, more research needs to be done to explore to what extent this trade-off could occur and whether it ultimately harms overall driving performance.

6. CONCLUSION AND FUTURE WORKS

We evaluated reaction time and accuracy of lane change test for verbal and nonverbal auditory cues manipulated along three dimensions (spatial, semantic, and temporal) in the presence of a visual cue. The results showed that the application of the multimodal displays could improve lane change test performance, but also showed that there are myriad interactions among variables.

Our results indicate that adding auditory cues could improve lane change test reaction time more than accuracy. The temporal dimension seems to be the most influential in performance. That is, preceding auditory cues improved reaction time. Spatially and semantically congruent auditory cues also facilitated reaction time. However, when these two dimensions conflict with each other, spatial congruency seems to have bigger impacts on performance. In other words, it is more difficult to ignore spatial location information than semantic verbal information just as in Barrow and Baldwin's research [1]. Moreover, when there is conflict between auditory cues and visual cues, having consistency in auditory cues would be more important than inconsistency within the auditory cue and partial consistency with the visual cue. In-vehicle technology designers will want to consider the plausible trade-offs when designing the multimodal warning or alert system.

MRT suggests that well-designed multimodal interfaces can allow drivers to more efficiently process information in distinct channels. Furthermore, MRT can readily account for the results of the current experiment. However, MRT includes only verbal information processing regarding auditory modality. The empirical evidence of the present study using non-verbal auditory cues supports the necessity of updating the model [24]. Then, the model will be able to better explain and predict the effects of non-verbal auditory displays of the multimodal interfaces. The results also showed sound's strong arousal effect, which can be better explained by the auditory preemption theory [25]. Certainly, more research is required to disentangle the various influences of auditory cues.

In future studies, it would be interesting to see the effects of the visual secondary task to increase driver workload. Given that Posner's experiment using the 200 ms interval was not conducted in the driving domain, more asynchronous intervals can also be tested in the experiment to see if there is any different threshold in multimodal perception while driving. More research on the definition of a reaction timer will be helpful in the maneuver level driving task compared with the operational level (go/no-go) driving task. We also plan to conduct a similar study using a higher fidelity simulator, which provides a more realistic driving environment.

7. REFERENCES

[1] J. H. Barrow and C. L. Baldwin, "Verbal-spatial cue conflict: implications for the design of collision-avoidance warning systems," in *Proceedings of the... international driving symposium on human factors in driver assessment, training and vehicle design*, 2009, vol. 5, pp. 405–411.

[2] C. D. Wickens, "Multiple resources and mental workload," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 50, no. 3, pp. 449–455, 2008.

[3] C. D. Wickens, S. J. Mountford, and W. Schreiner, "Multiple resources, task-hemispheric integrity, and individual differences in time-sharing," *Hum. Factors*, vol. 23, no. 2, pp. 211–229, 1981.

[4] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, Dec. 1976.

[5] W. Giang, E. Masnavi, and C. M. Burns, "Perceptions of Temporal Synchrony in Multimodal Displays," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2011, vol. 55, no. 1, pp. 1165–1169.

[6] C. Spence, "Crossmodal spatial attention," *Ann. N. Y. Acad. Sci.*, vol. 1191, no. 1, pp. 182–200, 2010.

[7] R. W. Proctor, H. Z. Tan, K.-P. L. Vu, R. Gray, and C. Spence, "Implications of compatibility and cuing effects for multimodal interfaces," in *Proceedings of the HCI International 2005*, 2005, vol. 11.

[8] J. D. Olsheski, "The role of synesthetic correspondence in intersensory binding: investigating an unrecognized confound in multimodal perception research." Georgia Institute of Technology, 2014.

[9] M. I. Posner, R. Klein, J. Summers, and S. Buggie, "On the selection of signals," *Mem. Cognit.*, vol. 1, no. 1, pp. 2–12, 1973.

[10] S. Sinnett, S. Soto-Faraco, and C. Spence, "The co-occurrence of multisensory competition and facilitation," *Acta Psychol. (Amst).*, vol. 128, no. 1, pp. 153–161, 2008.

[11] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman, *Engineering psychology & human performance*. Psychology Press, 2015.

[12] C. Wickens, J. Prinet, S. Hutchins, N. Sarter, and A. Sebok, "Auditory-Visual Redundancy in Vehicle Control Interruptions Two Meta-analyses," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2011, vol. 55, no. 1, pp. 1155–1159.

[13] J. M. Pieters, "Ear asymmetry in an auditory spatial Stroop task as a function of handedness," *Cortex*, vol. 17, no. 3, pp. 369–379, 1981.

[14] A. R. Mayer and D. S. Kosson, "The effects of auditory and visual linguistic distractors on target localization.," *Neuropsychology*, vol. 18, no. 2, p. 248, 2004.

[15] J. H. Barrow and C. L. Baldwin, "Semantic versus Spatial Audio Cues: Is There a Downside to Semantic Cueing?," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2009, vol. 53, no. 17, pp. 1071–1075.

[16] J. H. Barrow and C. L. Baldwin, "Individual differences in verbal-spatial conflict in rapid spatial-orientation tasks," *Hum. Factors J. Hum. Factors*

Ergon. Soc., p. 18720814553792, 2014.

[17] A. H. S. Chan and C. K. L. Or, "A comparison of semantic and spatial stimulus-response compatibility effects for human-machine interface design," *Eur. J. Ind. Eng.*, vol. 6, no. 5, pp. 629–643, 2012.

[18] S. Mattes, "The lane-change-task as a tool for driver distraction evaluation," *Qual. Work Prod. Enterp. Futur.*, vol. 2003, p. 57, 2003.

[19] ISO, "26022: 2010 Road vehicles–Ergonomic aspects of transport information and control systems–Simulated lane change test to assess in-vehicle secondary task demand," *Int. Organ. Stand.*, 2010.

[20] H. Tattegrain, M.-P. Bruyas, and N. Karmann, "Comparison between adaptive and basic model metrics in lane change test to assess in-vehicle secondary task demand," in *21st esv conference*, 2009.

[21] T. D. Wickens and G. Keppel, "Design and analysis: A researcher's handbook." Englewood Cliffs, NJ: Prentice-Hall, 2004.

[22] C. Spence and J. Driver, *Crossmodal space and crossmodal attention*. Oxford University Press, 2004.

[23] C. Koppen and C. Spence, "Audiovisual asynchrony modulates the Colavita visual dominance effect," *Brain Res.*, vol. 1186, pp. 224–232, 2007.

[24] M. Jeon, "How Is Nonverbal Auditory Information Processed? Revisiting Existing Models and Proposing a Preliminary Model," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2016, vol. 60, no. 1, pp. 1529–1533.

[25] C. D. Wickens, S. R. Dixon, and B. Seppelt, "Auditory preemption versus multiple resources: Who wins in interruption management?," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2005, vol. 49, no. 3, pp. 463–466.

SOUNDING OBJECTS: AN OVERVIEW TOWARDS SOUND METHODS AND TECHNIQUES TO EXPLORE SOUND WITHIN A DESIGN PROCESS.

Marcelo Pedruzzi Ferranti

Electronic Art Laboratory, Department of Art & Design
Pontífica Universidade Católica do Rio de Janeiro - Brazil
marcelopferranti@gmail.com

Rejane Spitz

Electronic Art Laboratory, Department of Art & Design
Pontífica Universidade Católica do Rio de Janeiro - Brazil
rejane@puc-rio.br

ABSTRACT

Sound is a neglected subject of today's products and services. The new technologies changed the way we interact with the people, objects and the world around us, thus, designers should aim at all senses, contemplating a multi sensorial experience. In this scenario sound becomes an important aspect to be considered during the project phase in a design process. Sound becomes part of the product identity and expression, the way the product talks to us. To foster this scenario designers should be aware of the possibilities and attributes of sound and how to explore them in a creative way. In this short paper we investigated published articles, workshops and publications to collect sound methods and techniques to be used into a design process. As a result, we proposed twenty essential sound methods that could be applied in a design thinking context. This is an ongoing research, part of a thesis experiment, since further methods and refinement could be added in the future.

KEYWORDS

Sonic Interaction Design, Sonification, Prototyping, Sound Methods, Auditory Display.

1. INTRODUCTION

We are living in the age of experiences. New technologies are changing the way we interact with others and with the world around us through products and services. The omnipresence of the internet, the connected objects (also know as internet of things, IOTs) and the huge adoption of smartphones created a whole new range of opportunities for designing products. At the same time with more technology, knowledge and capability in the hands of the designer, more complexity and variables take place in form of craft techniques, privacy and ethics. Current interactive products, services, and environments are appraised by their sensory attributes, in addition to their form and function [1]. Unfortunately, even with all this new range of possibilities, designers only focus on the visual aesthetics of products.

This is a consequence of the dichotomy of form and function stipulated by The Bauhaus school of thinking. However, products and services of today demand an holistic multissensorial approach. Designers should design for the senses. The products of everyday are part of our narrative, they are themselves part of a diegetic experience trough micro-narratives and micro-interactions with objects and the world around us. Hug [2] reinforces that computing technologies turn everyday artifacts into narrative, procedural objects. These objects consist in socio-cultural components in the narratives of our everyday lives, expressing our personality, status, emotions and attitudes. Thus, how to think in ways that can contemplate all the sensory aspect of a product. Since sound is the "voice" of things and relate to the manifestation of life, how to create expressive objects that fit in your day to day diegetic narrative?. Thus, this literally "disappearing" technology offers several opportunities for using sound in its design [2]. Opportunities to make sonically-enhanced products through design. The big issue is that designers are often unaware of the auditory domain, of its complexity and potential and largely ignore sound processing and synthesis methods. Sound can have a profound effect on the experience an emotional appraisals of everyday products in use [3]. How can designers embrace and understand sound in this age of narratives and experiences?

2. THE SOUND AESTHETICS

What is sound? In his book "Acoustic territories", the artist Brandon LaBelle investigated the role that sound can play in our culture, listening and contemporary society. Sound and auditory experience forms a primary sensual matter in continual contact with the body. The sonority of daily life is a deeply impressionable sensing, impinging on thought and feeling in ways that give accent to the shifting self. The physicality of sound, as a movement of air pressure, of vibration, of interpenetrating exchanges from all around, forms an enveloping and effective influence. Such experience fills everyday life with an ongoing material flux, forming a phenomenal life-force existing here and there in which we are deeply involved. [4]

There is a potential to convey numerous different messages by non-verbal sounds [5]. Although verbal portrayals and metaphors can be used (like "should sound like a kicking ball" or muffed, like "an old radio"), it is almost impossible to accurately describe nuanced modifications to sound, or to explain the full range of sounds one has access to through imagination [6]. However it's interesting to mention that sound and colors have a few sensory properties in common: sound is auditory and has properties such as tempo, pitch, timbre, and rhythm. Color is visual and has the

https://doi.org/10.21785/icad2017.021

properties of lightness, vividness, and hue. Although both sound and color map aspects of emotion, when used together there is a valuable way to convey meaning and information, forging sound and a visual stimulus that occurs at the same time into an "augmented perception". Chion [17] called it synchresis, an acronym formed by telescoping together the two words synchronism and synthesis. Thus, how can we make expressive artifacts, guided by narratives and metaphors. How can we make objects talk?

2.1 The Sonic Interaction Design

Sonic Interaction Design (SID) is a field that is positioned at the intersection of auditory display, ubiquitous computing, interaction design and interactive arts [7]. The author also reinforces that SID can be used to describe practice and inquiry into the various roles that sound may play in the interaction loop between users and products, services and environments. In this sense, SID follows the trends of the so-called third wave of human– computer interaction, where culture, emotion, and experience, rather than solely function and efficiency, are the scope of interaction between human and machines. Sonic Interaction Design algo study methods and techniques to think, explore and attach sounds into everyday objects. To the date, concepts and guidelines in relation to notification and warning sounds, such earcons, auditory icons and sonification strategies for representation of data through sound have already been widely discussed [8] [12].

Also the relationship between the body movement, gesture and their relations to sound [9]. Even software frameworks for prototyping sounds using parameters and controls made using MAX/MSP are proposed [7]. An exploration of basic sound design methods, which are inspired by methods of the Bauhaus, can be found in the publications of Franinovic and others [10]. This is a challenge, as the criteria for designing interactive sounds are different from the criteria that drive aesthetics (visual and material). As a consequence, designers tend to stick to seemingly "save" strategies, like simple signals or based on existing sound libraries [2].

2.2 The Sonic Methods And Techniques

2.2.1 Sounding objects should be "sounds of novel".

Sound has to have a character and identity of its own, rather than sonic references to familiar "natural" sounds. Interactive products should be conceived as completely new sounds, without too obvious references to a "heritage artifact".

2.2.2 Sounding objects should be expressive and emotional.

The relationship between sound and signified (iconic or indexical relationship) as found in auditory icons and earcons are insufficient. Meaning and information cannot be conveyed in a satisfactory way by simply mapping isolated parameters like pitch, volume, speed or timbre. To achieve expressive and emotional sounds, designer should overlap different sounds to create a sound polyphony.

2.2.3 Play with Sound Aesthetics (8-bit Sounds/ Cartoonification).

Try to emulate sound using onomatopoeias, distortion, echo, delays, reverb and add some effects to them.

2.2.4 Play with natural and rational sounds of everyday objects.

They can be useful as a starting point of a "soundstorming". Sound is closely related to physical, material processes and plays a core role in communicating "hidden" qualities of an object, like stability, solidity, etc. [2].

2.2.5 Transcode objects, images, movements and data into sound.

Transcode colors, shapes, objects, images, event buildings into sound using softwares that can convert specific parameters into the sound realm. Tanaka [9] did some experiments converting photographic images into sound.

2.2.6 Foley Techniques.

A classic technique from sound design applied in cinema. Try to reinforce the sound aspects of the elements using materials that are not necessarily the ones that you are trying to represent.

2.2.7 Digital frameworks using MAX/MSP and Pure Data (PD).

Digital tools that allow that physical sound models can be created and changed dynamically according to input from the user (via some kind of sensor or MIDI controller). Rocchesso and his colleagues [7] created a tool that allows the designer to experiment with sounds.

2.2.8 Sound Scenarios/Sound experience map.

Sound elements are injected into a recorded narrative of intended use cases. Try to create scenarios of use of the object considering every interaction with the object, people and the environment around you, using only sound. Also it's important to map the emotional reaction to them, something similar to the experience maps methods used in service design. Storytelling is a important catalyst for making sound design decisions [2].

2.2.9 Soundmarks.

Based on the work of Murray Schafer. Schafer states that sound that are strongly representative of a soundscape, like The Big Ben Bell resembles London or the foghorn and doves that resembles a port. Another example is the choro and samba that are tied with the Rio de Janeiro image. This concept of soundmarks can be applied to a town, a building, a country, or even to a season [11].

2.2.10 Collect Everyday Interactions.

A way of gathering interesting sounds around you in the day to day using a simple recorder or even your smartphone. They are important to create a "soundboard".

2.2.11 Soundboards/Sound collage.

Why not boost the traditional moodboard (a board full os images, quotes and scenarios that are intended to convey meaning) with sound? That could be sound from digital libraries, sound collected on the streets, sounds made using foley techniques and so on.

2.2.12 The "Sonic Incident".

Remember an incident in the last two days, where the sound was memorable, and concentrate on emotions felt (frustrating, surprising, fear, etc.) and sketch the situation alongside with the sound descriptions.

2.2.13 Sound personality/Sound persona.

How do you want the personality of your object would be? Evil? In a hurry? Happy? Cute? A child? An old man?

2.2.14 Acoustic Ecology.

Another principle based on the work of Murray Schafer. Acoustic ecology is the whole set of sounds that are heard during the experience with a product. It is important to design all sounds as an "ecology" as sounds have relationships that need to be managed appropriately in a holistic way to have a cohesive experience.

2.2.15 Vocal Sketching.

Vocal sketching involves the use of the voice, along with the body, to demonstrate the relationship between action and sonic feedback. Vocal sketching, in essence, is as simple and straightforward as it sounds: the designer uses his or her voice to produce the sound that would be generated in the sonic interaction. It can be used both for idea exploration and as a way to refine sound ideas.

2.2.16 Sonic Guessing Game.

Try to illustrate everyday situations or the scenarios using only sound. It serves very well as warming up for design sessions. [3]

2.2.17 Sonic Superpowers.

Imagine an object that could emit any type of sound that you could imagine of. For example, how to enrich the experience of using a microwave through sound? Imagine if that microwave could emit sounds of something being electrocuted, or maybe a looping drum machine that beats accordingly to the way the food spins?

2.2.18 Sound Walks.

Sound walks were originally proposed by Murray Schafer as an empirical methodology. When performing a sound walk, people are asked to navigate in a delimited area with open ears, remembering all the sounds heard. This exercise is an ear opener and a good starting point to enhance the understanding of the sound ecology around us.

2.2.19 Create an Audio Drama using sound libraries and Foley.

This is achieved by designing a short audio drama involving the collection of content, either from existing sound libraries or generated by the designers themselves. The analysis of the performance makes designers aware of how sounds affect gestures and how, conversely, gestures may affect the mental representations elicited by sound. [3]

2.2.20 Sonic Wizard of Oz.

It's a form of video prototyping in which an interaction is filmed and the sonic elements are added over the footage at a later stage, creating a video of a fake sonic interaction. Sonic overlaying gives the designer the best possible conditions for creating the desired sound. At the editing table, the variety of sonic materials available to the designer can be found or created, be they voice, everyday objects, music, downloaded sound samples, and the like.

3. CONCLUSION AND FUTURE STUDIES

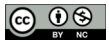
For the future work we plan to collect more sound methods and proposed new ones based on classic methods of design. This would complement the methods and techniques already collected in that article. We also intend to cross some thoughts and ideas of Pierre Schaeffer [13], Murray Schefer [14] [15], Elif Özcan [16] and Michel Chion [17] to structure a sound taxonomy which will serve as framework to explore new methods to think and understand sound in a design context. Finally, it would be crucial to validate these sound design methods in a real design project environment to test it's effectiveness and problems encountered in a real problem solving setting. These will be the basis of my ongoing master thesis.

Despite numerous advantages and ways to shape experiences in products and services, the sonic aspects are extremely neglected in the design processes. In this article we tried to understand the possibilities of tools and methods for sound in design. In the first moment, one important part was to understand the double-diamond and human-centered design approaches and all the dynamics of divergence and convergence thinking during the design process. And most importantly, to know some of the most used design methods and tools in those environments. In the second moment, gathering and research about sound methods was extremely valuable. These methods are pulverized in different articles by different authors, most of these articles are about workshops where these sound methods were tested and prototyped. Thus, it was possible to have a broader and deeper understanding of sound methods in design. The exercise of researching, mapping, collecting and organizing some of the methods for thinking sound in design processes was proven to be extremely valuable. At last, it was possible to have a detailed understanding about the state of the art of these tools and how they can be evolved, or even suggest new tools and techniques. It is important to understand that sound has an extreme ability to enrich products and services and there is a lot of unexplored potential. With new technologies and new possibilities there is still much to be done. A set of solidified methods is needed to think and design the "talking objects" of tomorrow.

BIBLIOGRAPHY

1. Serafin, S., Hoby, M. and Sarde, J. 2015. Product Sound Design: Form, Function and Experience. Proceedings from AM15, October 07-09, Thessaloniki, Greece.
2. Hug, D. and Kemper, M. From foley to function: a pedagogical approach to sound design for novel interactions. Proceedings from Journal of Sonic Studies, volume 6, nr, 1. January, 2014.
3. Rocchesso, D., Serafin, S., Rinott, M., Pedagogical approaches and methods. In Sonic Interaction Design,

MIT Press. Cambridge. Edited by Karmen Franinović and Stefania Serafin, 2013.
4. La Belle, Brandon. Acoustic Territories: Sound Culture and Everyday Life. New York: Basic Books, 2014.
5. Suied, C., Susini, P., Misdariis, N., Langlois, S., Smith, B.K. and McAdams, S. Toward a sound design methodology: application to electronic automotive sounds. Proceedings of ICAD 05 - Eleventh meeting of the international conference on auditory display. Limerick, Ireland, July 6 - 9, 2005.
6. Ekman, I. and Michal, R. 2010. Using vocal sketching for designing sonic interactions. Proceedings from DIS 2010. August 16 - 20. Aarhus Denmark.
7. Rocchesso, D., Serafin, S., Behrendt, F., Bernardini, N., Bresin, R., Eckel, G., Franinovic, K. Hermann, T., Pauletto, S., Susini, P., and Visell, Y., 2008. Sonic Interaction Design: Sound, Information and Experience. Proceedings from CHI 2008. April 5 - 10, Florence, Italy.
8. Buxton, Bill - Sketching User Experiences: Getting the design right and the right design. Morgan Kaufmann, 2007.
9. Tanaka, A., Bau, O. and Mackay, W. The A20: Interactive Instrument Techniques for Sonic Design Exploration. In Sonic Interaction Design, MIT Press. Cambridge. Edited by Karmen Franinović and Stefania Serafin, 2013.
10. Franinovic, K., GAYE, L., and Behrendt, F. Exploring sonic interaction with artifacts in everyday contexts. Proceedings of 14th International conference on auditory display, Paris, France, 2008.
11. Susini, P., Talotte, C., Misdariis, N., Dubois, F., and Carron, M., Designing sound identity: providing new communication tools for building brands “corporate sound”. Proceedings from AM’14, October 1 - 3, 2014, Aalborg, Denmark.
12. Gaver, W. 1989. The Sonicfinder: And interface that uses auditory icons. Human-Computer Interaction 4, 67-94.
13. Schaeffer, P. 1966. Traité des objects musicaux. Editions du Seuil, Paris, France.
14. Schafer, R. Murray. O ouvido pensante. São Paulo: UNESP, 1991.
15. Schafer, R. M. 1977. The tuning of the world. New York, 1977.
16. Özcan, E. Product Sounds: fundamentals and application. Doctoral thesis. TUDelft, 2008.
17. Chion, M. Audio-Vision: Sound on Screen. Columbia University Press, 1994.



PROGRESS TOWARD SONIFYING NAPOLEON'S MARCH AND FLUID FLOW SIMULATIONS THROUGH BINAURAL HORIZONS

Dr. Peter W. Coppin
Perceptual Artifacts Lab, Faculty of Design and Inclusive Design, Ontario College of Art and Design University, Toronto, Canada
pcoppin@faculty.ocadu.ca

Richard C. Windeyer
Centre for Drama, Theatre and Performance Studies, Knowledge Media Design Institute, University of Toronto and Perceptual Artifacts Lab, Ontario College of Art and Design University, Toronto, Canada
r.windeyer@mail.utoronto.ca

Daniel E. MacDonald
Biomedical Simulation Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, Canada
demacdo@mie.utoronto.ca

Dr. David A. Steinman
Biomedical Simulation Laboratory, Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, Canada
steinman@mie.utoronto.ca

ABSTRACT

Cross-modal data analytics—that can be rendered for experience through vision, hearing, and touch—poses a fundamental challenge to designers. Non-linguistic sonification is a well-researched means for non-visual pattern recognition but higher density datasets pose a challenge. Because human hearing is optimized for detecting locations on a horizontal plane, our approach recruits this optimization by employing an immersive binaural horizontal plane using auditory icons. Two case studies demonstrate our approach: A sonic translation of a map and a sonic translation of a computational fluid dynamics simulation.

1. CROSS-MODAL DIAGRAMS

Figure 1a shows some physical material that has been configured in a manner that represents, models, or maps a situation. Carved marks, as notches, are arranged on the material (wood) in geometric, topological, or iconic relations with other marks (notches). The result models, maps, or represents the concrete structure (or shape) of a geographic region.

Recruiting Larkin and Simon's classic definitions [1], this physical arrangement of notches is essentially a *diagram*, defined as items indexed to a plane. It is certainly not a sentence—defined as items indexed to a sequential list, where each item of the list is only adjacent to the item before and after it on the list—but it also does not rely solely on visual perception. It represents, or models, a situation both visually and tactilely, suggesting cross-modal mappings that transcend visual approaches that dominate contemporary digital media and data analytics research.

Taking this further, there is nothing inherently "visual" about any diagram. As noted, the spatial, topological and geometric properties of a diagram can be mapped to spatial properties of a tactile surface. In addition, they can also be mapped to spatial properties of sound. The items that are indexed to the spatial properties of sound can then be made available to perception via text-to-speech labels, earcons, or auditory icons.

This research was funded by the Canadian Natural Sciences and Engineering Research Council, the Social Sciences and Humanities Research Council, and the Heart & Stroke Foundation.

https://doi.org/10.21785/icad2017.042

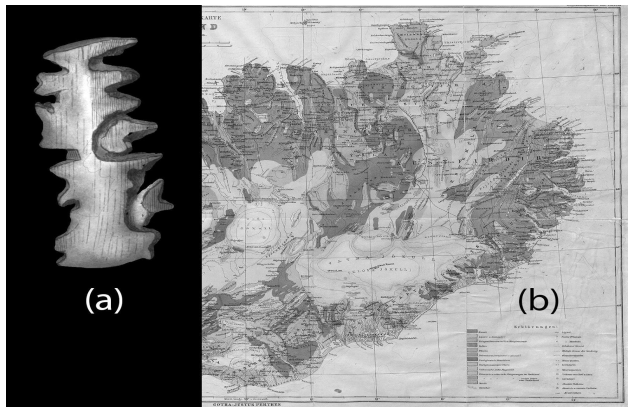


Figure 1: An indigenous diagram, which can be perceived both visually and tactilely (a); in comparison to one of European origin, which is purely visual (b). Adapted from [2, p. 250] and [3]. Public domain.

Although such a mapping from visual to sonic diagrams seems plausible, in practice, obstacles abound, particularly when attempting to construct high-density diagrams—a fundamental challenge faced in virtually all areas of information design. It is in this area of high-density sonification that this paper makes several contributions.

First, we report on our progress toward developing sonic translations of high-density visualizations that are currently accessible only through visual perception through two "grand design challenges": Charles Minard's classic infographic *Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813* (1869) [4][5, p. 40] (Figure 2), and simultaneously, aspects of computational fluid dynamics (CFD) blood flow data in order to augment conventional flow visualizations and increase ease of interpretation of the data.

In addition to reporting on our progress above, we describe how simultaneous development on two distinct design challenges reveal underlying principles and cross-compatible solutions (that underpin both examples), foreshadowing a provisional taxonomy of cross-modal mappings.

Finally, in the process of describing the reasoning behind our design choices, we provide an overview for how work from the ICAD community maps conventions that we are employing.

2. GRAND DESIGN CHALLENGE: SONIFYING AN EARLY STATISTICAL GRAPHIC

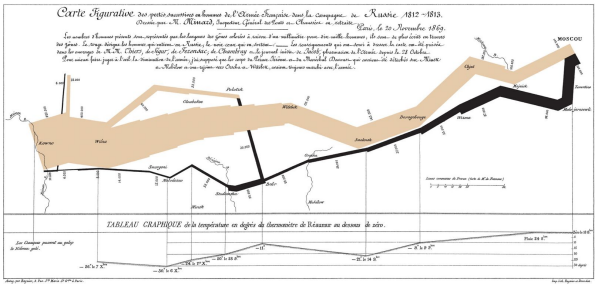


Figure 2: Charles Minard's *Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812–1813* (1869). Public domain.

Charles Minard's 1869 infographic is remarkable for its ability to compress six types of data (number of troops; geographical distance; changing temperature; latitude and longitude; travel direction; and location-specific dates) into a two-dimensional representation [4]. The journey undertaken by Napoleon's troops can be decomposed into several variables: *Troop Quantity*, *Troop Position*, *Troop Direction*, and *Troop Distance Travelled*. A visually perceived line on the rectangular map surface conveyed these variables: The positions of each point of the line conveyed latitude and longitude information (*Troop Position*), travel direction (*Troop Direction*), and geographical distance (*Troop Distance*). A legend in the map used written graphics to show that the line represented the troops' journey. Minard used line thickness to convey *Troop Quantity* throughout the journey.

2.1. Selection and design of auditory icons

Our sonic translation [7] employed sound recordings of marching footsteps (on various outdoor surfaces), seasonal environmental cues (such as birdsong, wind, and snow) to convey changes in terrain and season encountered by the troops, and battle sound effects (such as gunshots and cannon fire to indicate battle locations). The temperature fluctuations that occurred during the army's return were depicted by a tuning frequency of a sine tone.

2.2 Mapping auditory icons to a horizontal ground plane

Through the use of an existing dataset extracted from Minard's infographic [8] and Ambisonic spatial encoding techniques [9] within the Max/MSP programming environment [10], our auditory icons were indexed to a horizontal ground plane or "soundmap" (Figure 3). From a fixed listening position at the centre of the map, the changing locations of troops at each point in the journey, their direction, and distance travelled (*Troop Position*, *Direction*, and *Distance*) are conveyed.

Troop Quantity was conveyed at each point in the journey by manipulating a combination of density, pitch/frequency, and tempo of the auditory icon: fewer troop quantities corresponded to the sound of fewer troops marching at a slower tempo and higher pitch, whereas more troops were conveyed by the sound of more troops marching at a faster tempo and lower pitch. *Troop Quantity* was mapped to multiple sonic parameters simultaneously in order to replicate a key visual feature of Minard's efficient infographic design: the colour-coded layering of the army's advance and retreat trajectories (gold and black respectively), on top of each other.

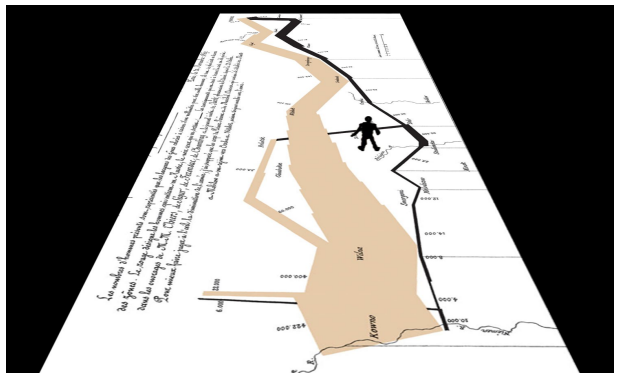


Figure 3: Sketch of the horizontal ground plane or "soundmap."

2.3 User interaction

While our initial design iterations supported only linear playback of the sonification, subsequent attempts at facilitating user interactivity produced several working metaphors which continue to inform our work—*World of Warcraft* and *Spot Mic Radar:*

1. *World of Warcraft* (WoW) evokes an immersive and navigable auditory experience of Minard's infographic, with or (ideally) without reliance on its visual presence.
2. *Spot Mic Radar* describes a more selective means of auditioning the sonification, whereby a user may rotate a virtual microphone around the horizontal ground plane from their fixed listening position at its centre. The user's speed of rotation affects the rate at which the changes in auditory icons may be discerned. At maximum speed, the effect is akin to the rotational detection of objects by a radar system. As with many real-world microphones, the directionality or "pickup pattern" of this "spot mic" is variable. This functions as a variable *cone of attention*, enabling a user to minimize cacophony by reducing the perceived width of the binaural image from fully binaural to monaural.

3. GRAND DESIGN CHALLENGE: SONIFYING ANEURYSMAL BLOOD FLOW

Just as Minard's infographic is remarkable for its ability to compress six types of data into a two-dimensional representation, the power of CFD visualization lies in how it recruits the parallel pattern recognition capability of vision to compress spatial and temporal qualities of numerous types of point data into a three-dimensional representation (two dimensions over time).

Our intention for this second design challenge is to enhance the analytic process of evaluating large datasets by augmenting visual representation through the application of the previously-mentioned sonification methods.

3.1. Context

Blood flow patterns are thought to contribute to the development of vascular disease [11][12]. Direct measurement of such flow features is challenged by the limited spatial and/or temporal resolution of current non-invasive medical imaging modalities. Instead, vascular geometries derived from non-invasive medical imaging are used as the basis for "patient-specific" CFD simulations of the blood flow dynamics [13].

The motivation of the Biomedical Simulation Lab for sonifying CFD datasets is to introduce a method by which clinicians can characterize recently-uncovered turbulent-like flows in intracranial aneurysms [14], in order to evaluate the likelihood of rupture. Inspired by the natural perceptual encoding between flow regime and sound, a model is proposed whereby the presence and properties of high-energy flow structures within the simulated aneurysmal flow will be communicated via sound.

3.2. Description of Data

These simulations result in data structures of high spatial and temporal resolution describing the system hemodynamics in three-dimensional space for a given period of flow [12]. Such results are conventionally inspected using the investigator's choice of interactive scientific visualization applications, such as ParaView [15]. Selection of perspective is required in order to prevent occlusion of important data while balancing the quantitative and qualitative properties of the flow.

By decomposing these volumetric aneurysmal flows in the frequency domain, the flow instabilities can be quantified [16]. Visually communicating these spatiotemporal instabilities ignores the existing strengths of the human auditory system for resolving frequencies as compared to visual perception of spatial frequencies. By sonifying these frequency-based operators, further spatiotemporal characteristics may be uncovered and will be used to *augment* conventional flow visualizations.

3.3. Existing techniques

Several methods of sonifying CFD-generated data have been explored to a limited extent over the last two decades. Klein and Staadt [17] provide methods for analyzing three-dimensional vector fields via sonification. Local flow direction, velocity, and vorticity are conveyed as the observer moves a sphere (the user's "head") through the vector field. Kasakevich et al. [18] furthered the work of Klein and Staadt by detailing methods for mapping flow properties of CFD data to a sonification model with the goal of increasing comprehension. Navigation occurs in a three-dimensional space where spatialization techniques are employed.

3.4. Toward a design solution: selection of auditory icons

Although our research began independently of the work of both Klein and Staadt and Kasakevich et al., it has converged upon a similar solution in which high-density data sonification is navigated within a three-dimensional auditory field.

Much like visual representation, sonification of this data may benefit from logical grouping; for example, by isolating turbulent-like structures using the methods of Khan et al. [14], pockets of high-energy flow instabilities can be inspected and interpreted much like looking at an isosurface in a visual representation. This requirement for grouping is explored in detail by Klein and Staadt. These volumetric structures may then be sonified based on the size, intensity, and motion. By *limiting auditory information within the perceptual field*, multiple high-energy pockets may be discerned under interactive inspection while interpreting local patterns and anomalies. The proposed auditory icon solution may resemble sonic forms akin to tornadoes swirling around and past the listener.

3.5. Prototype: a soundmap for user interaction

Prototyping began with a search for digital sound synthesis instrument capable of evoking a convincing impression of turbulence in fluid flow when driven by CFD data. Based on a computational model developed and documented by [19], our instrument employs frequency modulation (FM) synthesis in combination with envelope generators, bandpass filters, and physically-modelled tube resonators. Our instrument model was then ported to the SuperCollider programming language [20] in order to increase portability within the lab while easily interfacing with the existing file structures of our CFD data.

Our current prototype [21] situates the user within a three-dimensional virtual auditory space as an extension of our WoW paradigm (Figure 4). The auditory icon allows the local intensity of blood flow to naturally emerge from the velocity trace data. The user may explore a conventional visualization of the flow, where flow-inspired auditory icons are used to indicate the properties of high-energy flow structures. This approach to labelling remained throughout our iterative prototyping.

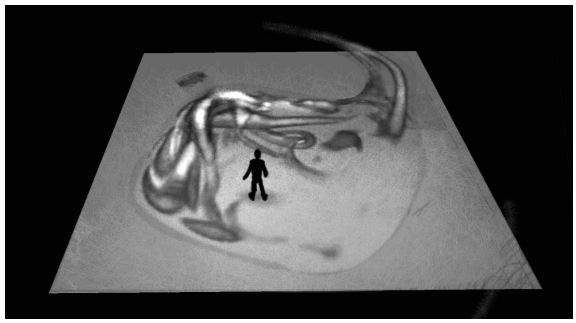


Figure 4: Sketch of a ground plane for representing blood flow trajectories within an aneurysm

4. DISCUSSION AND IMPLICATIONS

Let us now consider some underlying principles and mappings that transcend these two design challenges. These principles include i) the use of auditory icons for the labelling of items indexed to a plane and ii) the use of map conventions for limiting the perceptual field and reducing disorientation of the user.

Three techniques were employed in addressing the complexity of our two grand design challenges: the design of *auditory icons*, the *labelling of items indexed to a plane*, and the use of a variable *cone of attention* control to limit the user's perceptual field (our "spot mic radar" paradigm).

4.1. The labelling of indexed items by auditory icons

The design of auditory icons allows effective labelling of items indexed to a plane. To label items indexed to a plane in each diagram, auditory icons were positioned within a virtual spatial environment and conveyed via binaural audio.

For our sonic translation of Minard's graphic, we employed the binaurally perceived sound of troops marching along a trajectory as an auditory icon to convey *Troop Position, Direction* and *Distance*. *Troop Quantity* was conveyed by manipulating the auditory icon: fewer troops were conveyed by the sound of fewer troops marching, and vice versa.

Similarly, the intensity of local flow instabilities are conveyed using auditory icons designed to evoke impressions of the chaotic nature of turbulent flowing fluid. Regions of high-energy flow will drive the parameters of our synthesis model, creating a caricature of flow containing a broad range of sonic frequencies with more chaotic change in filter modulation.

4.2. Dealing with disorientation while reducing cacophony

The use of our "spot mic radar" avoids cacophony within a plane of indexed items and reduces disorientation. (*Cacophony* refers to the ambiguity caused by a number of cues impinging simultaneously.) Our sonic version of a retinotopic map—the World of Warcraft paradigm—avoided cacophony via a narrow detection sphere that avoided signals from both trips simultaneously.

Our blood flow sonification avoids cacophony by enabling the user to selectively choose their position within the flow while adjusting the size and scope of the perceivable auditory field. The use of auditory icons and logically grouped data allows several levels of detail to be presented simultaneously. The end-user design of our flow sonification differs from that of our Minard sonification; the augmentation of flow visualizations allows the user to orient themselves by way of the visual interface. Still, concepts such as the spot mic radar paradigm may be useful in communicating peripheral localization of high-energy flow structures.

5. CONCLUSION

This paper reports on two simultaneous grand design challenges with the intention of developing complementary approaches to the sonification of data produced by flow-like structures and dynamics—such as marching soldiers and blood flow. Both projects arrived at cross-compatible solutions that situate a single user within a navigable virtual auditory environment from a first-person perspective. Similarly, both required additional strategies for reducing cacophony by limiting the amount of sonic information present within the perceptual field. In addressing the challenges of sonifying fluid flow data, the natural perceptual encoding of flow regime is employed to convey the presence of unstable turbulent-like structures using auditory icons within a limited perceptual field. Further revisions of these prototypes will refine the proposed cross-modal taxonomy and techniques.

6. REFERENCES

[1] Larkin, J. H. & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–99. doi:10.1111/j.1551-6708.1987.tb00863.x

[2] Holm, G. & Garde, V. (1887). *Den danske Konebaads-Expedition til Grønlands Østkyst*. Retrieved March 6, 2017 from https://books.google.com/

[3] Thoroddsen, T. (1906). *Geologische Karte von Island* [Map]. Retrieved March 6, 2017 from http://islandskort.is/is/map/show/608

[4] Minard, C. J. (1869). *Carte figurative des pertes en hommes de l'Armée Française dans la campagne de Russie 1812–1813* [Infographic]. Retrieved May 15, 2017 from https://commons.wikimedia.org/wiki/File:Minard.png

[5] Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

[6] Corbett, J. (2001). Charles Joseph Minard: Mapping Napoleon's March, 1861. In D. Janelle (Ed.), *CSISS Classics*. Center for Spatially Integrated Social Science. Retrieved May 15, 2017 from http://escholarship.org/uc/item/4qj8h064

[7] Windeyer, R. C. (2017, May 10). *Sonification Prototype: a binaural translation of Charles Minard's infographic "Figurative Map of the Successive Losses in Men of the French Army in the Russian campaign 1812–1813" (1869)* [Video file]. Retrieved May 14, 2017 from https://vimeo.com/217423986

[8] Napoleon's March Data [Computer source code]. (n.d.). Retrieved May 15, 2017 from http://www.datavis.ca/gallery/minard/minwilk.sas

[9] Cycling '74. (n.d.). Max/MSP. Available May 15, 2017 at http://www.cycling74.com

[10] Morbiducci, U., Kok, A. M., Kwak, B. R., Stone, P. H., Steinman, D. A., & Wentzel, J. J. (2016). Atherosclerosis at arterial bifurcations: evidence for the role of haemodynamics and geometry. Thrombosis and Haemostasis, 115(3), 484–492.

[11] Turjman, A. S., Turjman, F., & Edelman, E. R. (2014). Role of fluid dynamics and inflammation in intracranial aneurysm formation. *Circulation, 129*(3):373–82.

[12] Taylor, C. A. & Steinman, D. A. (2010). Image-based modeling of blood flow and vessel wall dynamics: Applications, methods and future directions. *Annals of Biomedical Engineering, 38*(3), 1188–203. doi:10.1007/s10439-010-9901-0

[13] Valen-Sendstad, K. & Steinman, D. A. (2014). Mind the gap: Impact of computational fluid dynamics solution strategy on prediction of intracranial aneurysm hemodynamics and rupture status indicators. *American Journal of Neuroradiology, 35*, 536–543.

[14] Sandia Corporation & Kitware. (2005). ParaView [Computer software]. Retrieved May 15, 2017 from http://paraview.org

[15] Khan, M. O., et al. (2017). On the quantification and visualization of transient periodic instabilities in pulsatile flows. *Journal of Biomechanics, 52,* 179–182. doi:10.1016/j.jbiomech.2016.12.037

[16] Klein, E. & Staadt, O. G. (2004). Sonification of three-dimensional vector fields. In *Proceedings 12th SCS High Performance Computing Symposium (HPC 2004)*, Society for Modelling and Simulation International, 2004.

[17] Kasakevich, M., Boulanger, P., Bischof, W. F., & Garcia, M. (2007). Augmentation of visualisation using sonification: A case study in computational fluid dynamics. In IPT-EGVE Symposium, 2007. doi:10.2312/PE/VE2007Short/089-094

[18] Farnell, A. (2010). *Designing sound*. Cambridge, MA: MIT.

[19] McCartney, J. et al. (n.d.). SuperCollider [Computer software]. Retrieved May 15, 2017 from http://supercollider.github.io

[20] Windeyer, R. C. (2017, May 13) *Sonification Prototype: Blood flow caricatures in three-dimensional auditory space* [Video file]. Retrieved May 14, 2017 from https://vimeo.com/182031056

7. ACKNOWLEDGMENTS

The researchers wish to thank Dr. C. Chnafa and O. Omari (Biomedical Simulation Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto), D. Retren (Perceptual Artifacts Lab, Ontario College of Art and Design University), and A. Li (Perceptual Artifacts Lab, Ontario College of Art and Design University). Peter Coppin would especially like to thank his students R. Vroom and B. Biggs, as well as other students from INCD Lab 1 who worked on the "Sonic Wayfinding Project," for the many conversations that informed the development of the "World of Warcraft" paradigm.

CHALLENGES AND CONSTRAINTS OF USING AUDIO IN ONLINE MUSIC EDUCATION

Paulo R. A. Marins

Universidade de Brasília,
Departament of Music,
Campus Universitário Darcy Ribeiro, Prédio SG-2
CEP: 70.910-900 – Caixa Postal 4432
Asa Norte – Brasília-DF, Brazil
pramarins@gmail.com

ABSTRACT

Several online music courses have been developed lately by educational companies. In addition, many universities have been offering music online degree programs. Since these courses and programs are taught through distance education, many ICTs are used such as: recorded video, online software, social networks, and audio. Although audio is widely used in the online courses and degree programs that aim to teach applied music, only a few research reports have been published recently about this subject. This paper intends to clarify – through a literature review - some questions concerning this use and also aims to provide a discussion regarding the challenges and constraints of using audio in online applied music lessons. It is also hoped that the discussions made in this paper may lead to the development of research in the area of online music education as well as in the specific field of sound in learning.

1. INTRODUCTION

Online music instruction has been growing throughout the world at all levels, from elementary to higher education. In the US, for instance, there are 13 (thirteen) graduate and 1 (one) undergraduate online degree program in music that are NASM (National Association Schools of Music) accredited. In n Brazil, there are 3 undergraduate online degree programs in music education that are offered by public universities. Moreover, there are many other online resources such as Massive Open Online Courses (MOOCs) and Open Educational Resources (OER) that have courses on many aspects of music, such as music businesses, applied music, and so forth. This type of education is mediated by Information and Communication Technologies (ICT), and according to [1], although these technologies are widely used in the online courses and degree programs that aim to teach music online, only a few research reports have been published recently about this subject. Nonetheless, there have been have been many studies regarding the use of ICT in general teaching and learning processes online. [2] discuss the various technologies and media available for distance education. They state that different media can be used in the teaching and learning processes involved in online education. At the time they wrote the book (2007), printed materials (textbooks, manuals, course notes, and study guides) were the most common medium used in distance education. They also state that audio media it is underused and neglected by educators. Additionally, they emphasize that technology must be reliable and the video and sound quality must be good enough in order to not interfere with the message that is going to be delivered by the instructor. [2] also question the Internet2 and its availability, as according to them – up to date – this type of internet is not available in all schools and universities in the world. Another key statement that the authors make is that educators tend to stick to a particular technology and deliver all the content of their course using that particular ICT. They also say that no matter how well designed the content is, it does not mean that the students are going to use it. It can be inferred then that it is also important to acquire feedback from the students after designing a course. Moreover, regarding the specific use of some digital technologies, [3] for instance, states that audio activates more than one sensory channel and consequently contemplates different profile of learners. [3] also highlights the advantages of employing multimedia content as – according to the researcher - they can: help developing skills, allow multiple modalities of learning, enhance the interactivity, leave the student choose whether he wants to learn individually or not, allow a better comprehension and therefore can be helpful in all of the stages of the learning processes. [4] points that the main advantage of the use of ICT is that the distance, which is intrinsic to the distance education modality, can become proximity as applications such as audio and video conferencing can allow a simultaneous dialogue between teachers and students. However, if using audio in online music education can enhance the learning process – as the aforementioned researchers state – it is then useful to understand why audio being underused in this modality of education as well as the challenges and constraints of using this specific type of media in online music education.

2. AUDIO IN ONLINE MUSIC EDUCATION

Although underused – as stated by some researchers cited above - audio is normally employed in online music instruction in a number of ways, such as: recorded audio files in order to provide musical examples or a ‘library’; a recorded video lesson (audio in this case is usually recorded and edited

https://doi.org/10.21785/icad2017.043

separately), or in a web conference in which instructor and students can have a synchronous communication and audio of the conference is streamed through the web. In order to illustrate the underuse of audio in online music education it is worth mentioning the study of [5], who conducted a research about learning objects in the guitar course of the distance education undergraduate degree program in music of Universidade ode Brasília – UnB, Brazil. From more than 300 (three hundred) learning objects found in the course, only 5 (five) were audio files. This may be related to the fact that - to the author's knowledge - most of the studies concerning the use of audio in online music education are related to video or web conferencing. Additionally, [6] conducted a research about teaching guitar online in the distance education undergraduate degree program in music education of Universidade Federal do Rio Grande do Sul – UFRGS, Brazil. He concluded that – despite technical problems that might occur - videoconferencing is the most effective tool for teaching guitar master-classes online. This might explain the number of studies related to the use of videoconferencing for teaching music online as well as the fact that the research works about the use of audio in online music education is focused on videoconferencing. [7] analyzed the use of videoconferencing technology by investigating a college professor who taught applied music lessons to an eighth grade trumpet student. The researcher states that videoconferencing based lessons face lots of challenges in aspects such as visual limitations, body movement, audio and video delay, and sound volume control. 2 (two) of the problems listed by [7] (delay and volume control) are related to audio. [7] then concluded that videoconferencing based lessons although feasible still could not replace face-to-face lessons, but rather could be used as a supplement for music instruction. The researcher was not the only researcher to mention technology constraints in videoconferencing based music education. [8] investigated the viability of piano lessons through the use of videoconferencing program Skype. Due to the lack of Internet 2 in Zambia, where the study took place, the cited researchers affirmed that the students in their study had to download video content rather than stream. This might be related to what [1] affirms, when the researcher states that good quality videoconference is a recent technology and restricted to limited populations. It is also worth mentioning that videoconference software like the aforementioned Skype was not developed for broadcasting music but rather to convey speech. Thus the frequency range of the videoconference audio content is between 80 to 80.000 Hz - whereas it is known that the frequency range of the human hearing is approximately 20 to 20.000 Hz. Therefore – in an applied music lesson for instance - it might be difficult for the instructor to give feedback to their students, especially in terms of artistic and aesthetic aspects of music [9]. For this reason, this type of software cannot be considered as ideal for music lessons, although it can be very useful in terms of oral communication between the instructor and the student. Also, even with a good internet connection there is some audio compression involved in the process and this may interfere in a music lesson. According to [10] this compression can make the job of a music teacher difficult in terms of the perception of certain aspects of the performance of a student. For the reasons listed above, [10] affirms that videoconferencing software as Skype is not ideal for online music instruction. Another common issue in synchronous communication is the so-called latency. This occurs due to the fact that there is a number of data conversion steps involved in the audio stream process (the analog sound is captured by the microphone, converted to digital and then converted into a format that can be transmitted), and in each conversion step can generate a delay effect or the so-called encoding latency. Obviously that latency can interfere in the perception of the sound produced by an instrument of a student. For addressing this issue, the Conservatorio di Musica Giuseppi Tartini (Trieste,Italy) developed a piece of software called Low Latency (LOLA). An initial research conducted by [11] in order to investigate the potential of the referred piece of software in terms of online simultaneous music performance and distance learning. The experiment involved a classical masterclass, a jazz lesson, and an old-time fiddle session. The researchers compared LOLA to other videoconferencing software such as Skype and Polycom. The results indicated that LOLA seems to be more efficient than Skype and Polycom in terms of online music making, teaching and learning. However, [11] state that – although the software is free – LOLA requires specific equipment with estimated costs of $5.386 which can prevent some schools, universities and professionals to use the referred software. Moreover, [12], in a recent study in which the researcher investigated instructors and students' behaviors in online music lessons found that the quality of the audio and visual interaction is central to the achievement of instructional success.

3. INITIAL REMMARKS

Based on the above discussion, some initial remarks can be made concerning challenges and constraints about using audio in online music education.

Audio is underestimated and underused. Researcher such as [2] and [5] support this argument. This occurs despite the fact that [3] reinforces the effectiveness of audio in online music education.

Although audio can be used many ways for teaching music online such as: an audio recorded lesson; a recorded music example; an audio library; a play along file, and so forth, the research studies tend to focus on the use of audio in videoconferencing or synchronous applications. One possible explanation for this may lie on the fact that – according to [4] – synchronous ICT may turn the distance which is intrinsic to online music education into proximity. Thus the use of video and web conferencing and therefore the focus on audio used in these type of applications.

Standard videoconferencing software available are not developed for music education. Unfortunately, this can compromise the effectiveness of the music learning process. Nonetheless, new technologies are being developed [10] [11] and the results are promising. Thus, it can be inferred that this issue it is likely to be addressed in a near future.

There are still many technical constraints that may prevent instructors to use audio in online music education. Some of them are related to: Internet connection, the aforementioned inadequate software, latency, data compression, audio delay, volume control, and so forth. Although some of these can be minimized by the use of adequate equipment, it is worth mentioning that technology is still not accessible to everyone being restricted to limited populations [1].

4. FUTURE RESEARCH

Based on the literature and also on the discussion made above, some possible research topics can be elicited concerning audio in online music education.

Research about asynchronous audio, for instance, since to this author's knowledge there is little research about this topic and there are other possibilities of using audio in online music education rather than in video or web conferencing.

It would also be worth investigating the specificities of audio in videoconferencing. The studies discussed in this paper tend to focus on many aspects of the videoconferencing but not particularly on audio.

Music education researchers should also try to develop online free applications for online music education taken into aspect the specificities of music. Although the results of LOLA are promising, this piece of software requires equipment that may not be accessible for populations that live in poor areas.

This paper – through a discussion based on the literature – aimed bring some questions concerning the challenges and constraints of using audio in online music education. In addition, some initial remarks could be drawn from the literature researched and some possible research topics were elicited. It is hope that the discussion made here may be useful for researchers of the field and also may help in the development of the research in the area of distance music education.

5. REFERENCES

[1] J. Bowman, "Online Learning in Music**.** Foundations, Frameworks, and Practice*s"*. Oxford University Press : New York, 2014.

[2] M. Moore and G. Kearsley "Educação a distância: uma visão integrada". Tradução: Roberto Gelman. São Paulo: Cengage Learning. 2007.

[3] P. Fahy, P. "Media characteristics and online learning technology. In: Anderson, T and Elioumi, F., **"**Theory and Practice of Online Learning"Athabasca,cde.athabascau.ca/online_book, 2004.

[4] O. Peters, O. "Didática do Ensino a Distância". São Leopoldo, Editora Unisinos, 2001.

[5] P.R. L. Figueirôa, P. R. L. "Um Estudo sobre Objetos de Aprendizagem no Âmbito do Curso de Licenciatura em Música a Distância da UnB". Trabalho de Conclusão de Curso. Universidade de Brasília. 2016

[6] P.D.A Braga. "Oficina de violão: estrutura de ensino e padrões de interação em um curso coletivo a distância*"*.Tese de Doutorado em Música. Universidade Federal da Bahia. 2009.

[7] R. J. Dammers, "Utilizing internet-based videoconferencing for instrumental music lessons" . Update: Application of Research in Music Education, 28 (1). 17-24, 2009.

[8] N.B. Kruse, S.C. Harlos, R. M. Callaha, and M.L. Herring. "Skype Music Lessons in the Academ: Intersections of Music Education, Applied Music and Technology". In: Journal of Music, Technology & Education, v.6, nr1, pp 43-60, 2013.

[9] B.K. Sheppard, G. Howe, G, and T. Snook . "Internet 2 and Musical Applications". Proceedings of the National Association of Schools of Music 84th Annual Meeting**,** Seattle. 2008.

[10] D. Gohn. "Educação Musical com as tecnologias da EaD". In: Música e Educação: Série Diáologos com o Som. Editora da Universidade do Estado de Minas Gerais. Barbacena. 2015.

[11] H. Riley, R. Macleod, M. Libera. "Low Latency Audio Video Potentials for Collaborative Music Making Through Distance Learning". In: Update: Applications of Research in Music Education. Published online. DOI: 10.1177/8755123314554403. 2014.

[12] K. J. Dye, "Student and instructors behaviors in online music lessons: An exploratory study*"*. International Journal of Music Education, 34 (2): 161-170. 2016.

ECHOEXPLORER™: A GAME APP FOR UNDERSTANDING ECHOLOCATION AND LEARNING TO NAVIGATE USING ECHO CUES

Wenyu Wu, Rudina Morina, Arley Schenker, Alexander Gotsis, Harsha Chivukula, Madeline Gardner, Felix Liu, Spencer Barton, Steven Woyach, Bruno Sinopoli, Pulkit Grover and Laurie M. Heller

Carnegie Mellon University
Pittsburgh, Pennsylvania

ABSTRACT

Echolocation -- the ability to detect objects in space through the perception of echoes from these objects -- has been identified as a promising venue to help visually impaired individuals navigate within their environments. The interest is in part because a proof-of-concept exists: certain visually impaired individuals are able to navigate using active echolocation. Why, then is echolocation is not in more widespread use among visually impaired individuals? It is possible that a lack of systematic echolocation training platforms has impeded individuals in picking up this skill. We designed a game-application that serves as a training platform for individuals, sighted or not, to train themselves to echolocate. Preliminary testing from both sighted and visually impaired individuals showed that users uniformly understood the game, although their enjoyment of the game was mixed. Although a number of game features could be improved, it is a promising training tool prototype for individuals learning to use echo cues for navigational purposes.

1. INTRODUCTION

Visual impairment is a significant and growing problem. The need to address this medical problem has led to the development of many systems that attempt to at least partially restore the visual sense. Many sensory substitution systems have been implemented as navigational aids for visually impaired individuals, such as those that use tactile inputs [1] or sounds [2] to substitute for or supplement vision. Our approach in particular has been inspired from studies that have shown the ability of some blind individuals to exhibit auditory navigation skills using echolocation. There is some evidence that this may be, in part, due to cortical reorganization and recruitment of visual areas for auditory processing [3].

https://doi.org/10.21785/icad2017.040

Echolocation -- the ability to locate objects in one's surroundings by sensing the reflections of an emitted sound -- is a promising venue for conveying visual information and aiding navigation of visually impaired individuals. Echolocation includes both passive and active processes, via sensing existing sound reflections and emitting one's own referent sound, respectively.

This raises a profound scientific question: *Why, then is active echolocation not in more widespread use among visually impaired individuals?* It is possible that this is due to lack of training. Currently, echolocation training is typically limited to self-training by trial and error or laborious training with another echolocator. Some research has used auditory virtual reality to simulate acoustic environments for studying echolocation [4], [5]. It is possible that many more individuals could train themselves to echolocate if a systematic platform were available to enable them to do so.

This work focuses on developing a mobile phone game-application that attempts to train individuals, sighted or not, to echolocate. The goals of the application are manyfold. First, it will help us assess whether echolocation is a skill that is amenable to being learned by sighted and visually impaired adults. Second, it will help obtain the science for developing systematic techniques to train people to echolocate. For instance, the artificial environment of a game-app can be made free of the audio and echo clutter of the real world, as well as make use of *physically unrealistic* echoes that one would not encounter in the real world but are nevertheless useful for training purposes.

2. BACKGROUND ON ECHOLOCATION AND AUDITORY PERCEPTION

Conveying spatial information through echoes relies on a few important aspects: (1) the time between the

reference sound (such as a mouth click) and the echo, known as *echo delay*, (2) the difference between the echo timing for the right and left ear, known as the *interaural time difference* (ITD), (3) the difference between the sound levels entering the ears, known also as the *interaural level difference* (ILD), and (4) the angle-dependent alterations in the frequency spectrum produced by the head and external ears (pinnae) known as the head-related transfer function (HRTF). HRTFs help to externalize sounds presented over headphones and also provide some vertical location cues.

Echo delay is important in conveying object distance. If an object is further from the observer, the echo takes longer to reach the observer. Likewise, ITDs and ILDs are important in conveying direction information. If, for example, an object is located to the right of the observer, the echo will reach the right ear first [6]. Echo level is not as robust a cue to distance as echo delay because it varies not only with distance, but also with frequency and the absorption characteristics of the surrounding reflective surfaces.

3. GAME DESIGN AND IMPLEMENTATION

The goal of this application is to provide a training platform for learning echolocation. We designed a game that requires the user to navigate through various mazes using simulated echoes. An avatar is used to represent the current location of the user in the maze. At any time, the user can tap to instruct the application to play echoes based on the current location of the avatar within the maze. The application plays a pre-generated click followed by a realistic echo that conveys spatial information about the maze through echo delays, ITD, and/or ILDs. For instance, if the user is facing a close wall straight ahead, the referent click and the resulting echo will be heard in quick succession.

Mazes are carefully chosen such that specific navigation skills, such as learning when obstacles are straight ahead, to the left, or to the right of the avatar, can be learned in sequence. As the game progresses, the mazes and auditory cues become more sophisticated and thus more difficult to navigate. Throughout gameplay, the app collects data about the performance of the user, and these data are eventually transmitted in a secure, confidential manner to a server. This includes metrics like the number of times the user hits an obstacle and the time it takes the user to complete a particular maze. These data provide the foundation that will allow us to improve on the scientific understanding of echolocation.

To ensure user friendliness for visually impaired individuals, feedback from the Disability Services Center at Carnegie Mellon University as well as the Blind and Vision Rehabilitation Services of Pittsburgh was incorporated in design of the project, and will continue to inform our improvement of the app.

3.1 Programming environment

We used Unity, a cross-platform game engine which provided us with a software framework for the creation and development of the game [7]. Unity takes care of the majority of low-level infrastructure, so the developer can focus primarily on game design. Our current development prioritizes the Android version, but we expect the game to be available for iOS in the future.

Although Unity is well-suited to this project, it has some drawbacks. For example, Unity is incompatible with many smartphone accessibility applications, and its user interface has been designed for sighted users and developers. In addition, several of the control gestures used in the game are also meaningful in accessibility applications, which may be confusing. However, it was our conclusion that Unity's advantages far outweigh its disadvantages. Future testing will inform the best way to overcome the aforementioned drawbacks.

3.2 Maze design

The player's avatar (visually represented as an arrow) is inside a maze with walls and walkable areas (e.g., Figure 1). A green exit sign indicates the goal location that the avatar needs to reach to complete each level. This visual display is turned off in gameplay but can be helpful in debugging. The unit for player movement is one tile, and the entire maze is a 9 x 9 corridor-based grid system. The first few mazes that a user completes are simple to navigate, providing one novel echo type and/or requiring one novel navigational input (such as turning).

During this tutorial phase, users are given voice instructions drawing their attention to certain echo cues, providing hints after idling time or a crash into a wall, and familiarizing users with the gestures necessary for

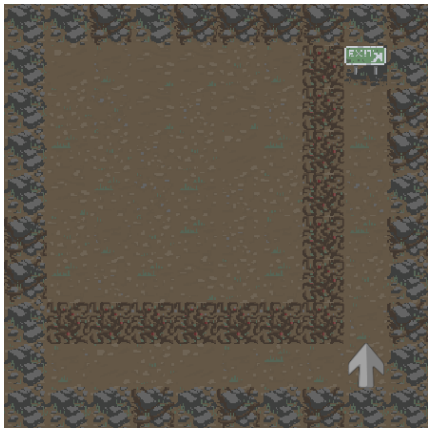


Figure 1. An example maze in the game, with two hallways. Player's avatar is indicated by the silver arrow. The level's exit is indicated by the green rectangle.

gameplay. This ensures that the user learns some basic navigational skills and begins to understand how echoes sound in simple scenarios. As the game progresses past the tutorial, the mazes become more sophisticated by using more complicated junctions and additional hallways. The game incorporates "structured randomness" to enable infinite gameplay levels.

3.3 Echo design and generation

Rather than generating echoes in real time, the game uses previously generated echoes that are loaded into the game for each physical situation that the avatar finds itself in. This avoids the computational overhead of generating sophisticated echoes in real time. Echo delay between the reference click onset and the first arriving echo was a reliable cue to the exact distance between the avatar and the wall ahead. In addition, we enhanced the sense of immersion in the game by using acoustic simulation and HRTFs. We used an individual HRTF [8]; however, combinations of HRTFs may have different acoustic properties than individual HRTFs and may be worth using in future versions [9]. Echoes varied according to distance from each of the surrounding walls and interactions at different types of hallway junctions so that users could distinguish left from right turns. These ambient cues to hallway location were updated every 2-3 blocks because of the many stored sounds required for all possible maze configurations. The exit location is represented as stairs, indicating that users must climb the stairs to get to the next level. The echoes generated adjacent to a set of stairs were judged to be sufficiently different from those adjacent only to walls, and were thus chosen to uniquely cue the exit location.

We used Odeon, a room acoustic simulation and measurement software, to artificially create echoes (Odeon A/S, Kongens Lyngby, Denmark). Users can input specific 3D room designs from which Odeon generates sounds using a ray tracing algorithm. We created several different hallways in Odeon to use in the game. In Odeon, we set the materials of the walls to 13mm plasterboard on frame, with a 100mm empty cavity. The floor and stairs materials were set to a wooden floor on joists. Within each of these rooms, a virtual cardioid sound source emitted a mouth click that had been previously recorded using a binaural microphone in our laboratory. The echoes produced by that click, but not the original click, were recorded from a location directly behind the sound source. We produced a set of echoes in response to this click for each tile location in the room. Using our own software written in Matlab™, these echoes were then placed at the appropriate time delay after the source click based on the distance from the front wall. We also used Matlab to further increase the loudness of each set of echoes to make the various echo cues obvious. Every 15 levels in the game, echo volume relative to initial click volume decreased by 2 dB to keep the game challenging. Further details regarding the creation and effectiveness of these sounds can be found in another paper in these proceedings [10].

3.4 Interface design and accessibility

Given the sophistication of the echoes in the game, users must wear headphones to play. Users navigate the game using a series of control gestures, which are introduced via a set of voice instructions in the tutorial phase of the game. Users have access to a menu, read vocally, through which they can play the tutorial, play the main game, or hear gesture instructions, and they can access this menu at any time. Within levels, the perspective of the game is first-person, so users swipe up to move one block forward no matter what compass direction they are facing. At each location, users tap the screen with one finger to hear the echo corresponding to that location. To turn right or left, users swipe in that direction. After swiping, to confirm their move, users hear a swoosh sound in the same ear as the direction they swiped. If they swiped up, users hear the sound in both ears. This sound is sufficiently different from the echoes. If a user crashes into a wall, they hear a crunch sound. When a user believes they have reached the exit location, denoted by a stairs echo, they use two fingers to tap once. If correct, users hear a congratulatory sound before starting the next level.

Though this game is intended for visually impaired individuals, a minimal amount of visual information is left for development purposes and to orient users with normal or partial vision. However, the game is designed to be played with all visuals turned off. This is expected to provide improved echolocation training by nearly equating the cues available to sighted and visually impaired individuals.

4. EVALUATION AND CONCLUSION

Survey data from 6 sighted participants using a beta version of the game indicated that it was initially well liked. After 15 hours of game play, participants were, on average, neutral about how fun it was and uniformly said that they would prefer not to continue playing the game. Data regarding the improvement of these individuals on the game and on a natural echolocation task are reported elsewhere in these proceedings [10]. All participants responded that they understood how to play the game and did not find the game frustrating. The control gestures were rated moderately easy to use.

Additionally, 8 visually impaired participants used an early version of the app for 5-10 minutes and provided survey data on their experience. On average, these users would play the game on their own (5.3 out of 7) and said it was fun (5.1 out of 7). They also provided qualitative feedback about that version of the game, which we took into consideration while making the current version. In combination with the ratings from the sighted users, these responses suggest that although the game could be made more engaging, it has promise as a convenient resource for supplemental echolocation training.

In conclusion, EchoExplorer™ is an app in development that has the potential to be an easy, accessible way for people to increase their awareness of echo cues. Future versions of the app may incorporate scoring or more complex obstacles to increase users' enjoyment of the game, as well as improve its compatibility with accessibility applications. If users find the game fun, they may use it for longer than they would a more boring training task. We hope to make future versions of this game useful to individuals who want to use echoes to be more aware of their environment for the purposes of navigation.

5. ACKNOWLEDGMENTS

We thank Sarah Kwan, Tejal Kudav, Kiran Matharu, Daya Lee, Jessica Kwon, Jacqueline Hon, Alan Lu, Aaron Steinfeld, Jeff Bigham, Chieko Asakawa, Catherine Getchell, Art Rizzino, and participants at the Pittsburgh Blind and Vision Rehabilitation Services, for helpful discussions, feedback, and pointers. We also thank Google Inc., the NSF Center for Science of Information (CSoI), CMU's SURG program, and the NSF REU program for their generous support.

6. REFERENCES

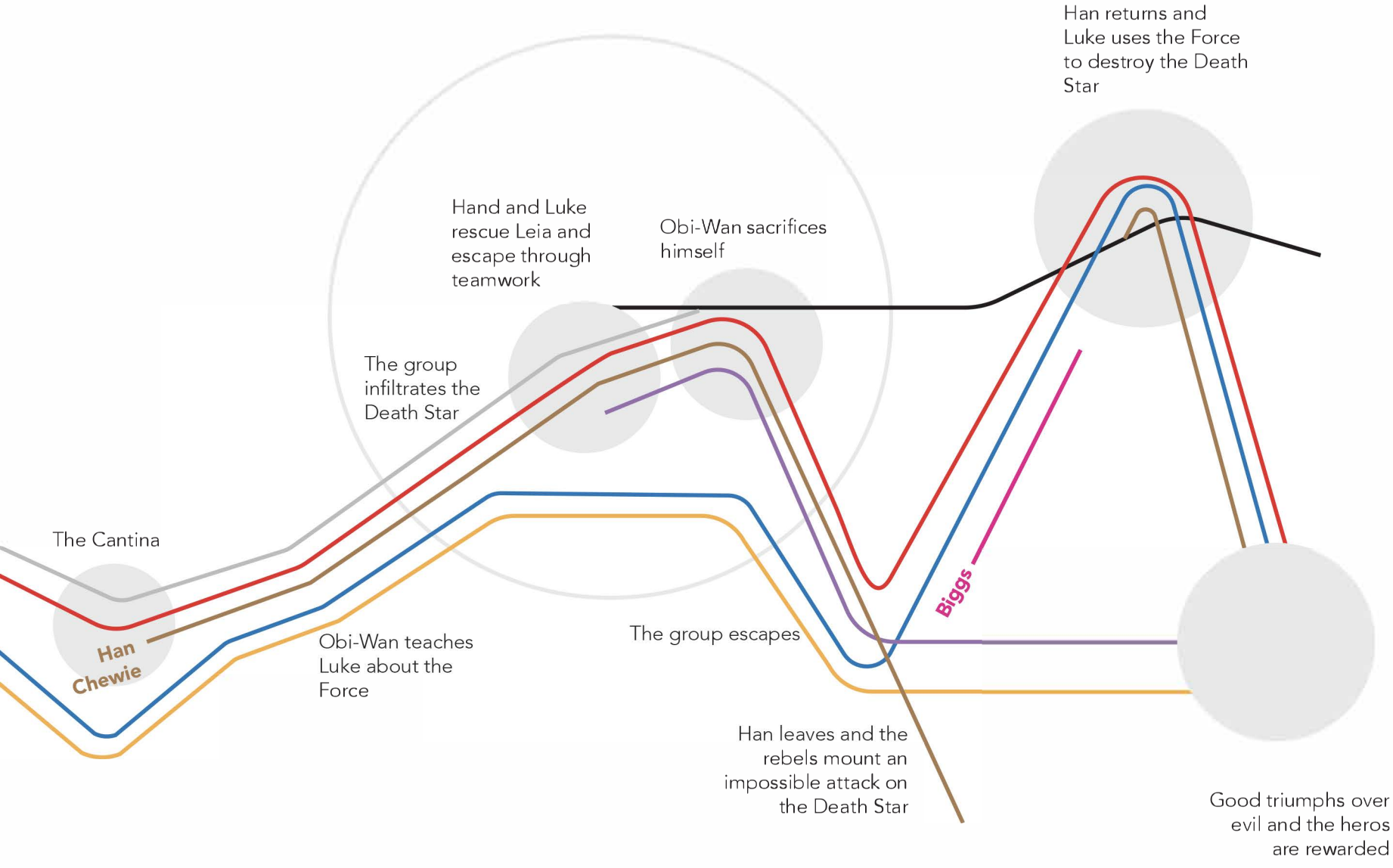
1. Y. Danilov, M. Tyler. Brainport: an alternative input to the brain. J Integr Neurosci. 2005;4: 537–550.
2. L. Kay. A sonar aid to enhance spatial perception of the blind: engineering design and evaluation. Radio and Electronic Engineer. 1974;44: 605.
3. O. Collignon, P. Voss, M. Lassonde, F. Lepore. Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. Exp Brain Res. 2009;192: 343–358.
4. B. F. G. Katz and L. Picinali, "Spatial audio applied to research with the Blind.," in *Advances in Sound Localization*, 2011, pp. 225–250.
5. D. Pelegrin-Garcia, M. Rychtáriková, C. Glorieux, and B. F. G. Katz, "Interactive auralization of self-generated oral sounds in virtual acoustic environments for research in human echolocation," in *Proceedings of Forum Acusticum 2014*, 2014.
6. J. Schnupp, I. Nelken, A. King. Auditory Neuroscience: Making Sense of Sound. MIT Press; 2011.
7. R. H. Creighton. Unity 3D Game Development by Example: A Seat-of-Your-Pants Manual for Building Fun, Groovy Little Games Quickly. Packt Publishing Ltd; 2010.
8. E. De Sena, N. Kaplanis, P. A. Naylor, T. van Waterschoot. Large-scale auralised sound localisation experiment. AES 60th International Conference 2016.
9. L. S. R. Simon, N. Zacharov, and B. F. G. Katz, "Perceptual attributes for the comparison of head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 140, no. 5, pp. 3623–3632, 2016.
10. L. M. Heller, A. Schenker, M. Gardner, F. Liu, "Evaluating two ways to train sensitivity to echoes," in *ICAD Proceedings 2017*, 2017.

INSTALLATIONS

Benjamin Andrew
beandrew@gmail.com
The Pennsylvania State University

Story Scores - an installation

This experiment creates original music based on story maps for popular films. Graphical depictions of plot, character arcs, and dramatic tension will be transposed to variables like frequency, amplitude, and changes in melody. Alongside these data-driven musical compositions, the process will be reversed to draw story maps based on existing musical pieces, exploring the structural shapes of music and fiction through translation.



SPAGHETTI OF STORMS: AN INSTALLATION OF TROPICAL STORM DATA SONIFICATIONS

Mark Ballora

School of Music/School of Theatre
The Pennsylvania State University
University Park, PA 16802 USA
ballora@psu.edu

Jenni Evans

College of Earth & Mineral Sciences
The Pennsylvania State University
University Park, PA 16802 USA
jle7@psu.edu

This will be an installation related to a poster shown at the 2015 ICAD, which described sonifications of tropical storms [1]. These sonifications are currently on exhibit at Penn State's Earth and Mineral Science Museum. A kiosk at the museum contains a video monitor and several sets of headphones. The video interface allows users to select from a group of 11 storms (Figure 1) by clicking on a button. When a storm is selected, a Google Earth map appears on one side of the screen that shows the path of the storm, along with a text box displaying the storm's name, dates, and countries affected. On the left, a satellite video of the storm plays, with a soundtrack that sonifies its properties (air pressure levels, latitude, longitude, and symmetry). Video control buttons allow the video to be played, paused, or stopped, and a slider allows users to "scrub" through the video at will.

A button on the right side of the screen allows users to display a mini tutorial window (Figure 2), where the sound mappings are explained and demonstrated.

An extension to this exhibit shows the "Spaghetti of Storms" from the year 2005 (Figure 3). A six-minute video, created by Wade Shumaker of the College of Information Sciences and Technology, will show the paths of some 150 storms that took place over the course of that year, with a sonification of each. (These sonifications are based on those of the exhibit, but modified due to the shorter playback time.) This additional video will be incorporated into the kiosk exhibit by the time of the ICAD 2017 conference.

The video/sound combinations present storm activity in an engaging and informative modality. Viewers have remarked that the satellite videos are far more engaging and visceral with the soundtrack added. There is also an added level of information provided by the sonification: the meteorologist who commissioned the videos found that some information about air pressure was sometimes not detectable from the video, but could be heard readily through the sonifications. The new video that gives the year's summary of storms has the potential to being a valuable pedagogical tool. Storms are typically thought of as singular, catastrophic events. Seeing individual storms as elements of a larger pattern will create a new perspective on global storm activity.

[1] Ballora, Mark. "Two examples of sonification for viewer engagement: Hurricanes and squirrel hibernation cycles." Extended abstract, *Proceedings of The 21st International Conference of Auditory Display (ICAD 2015)*, July 8-10, 2015, Graz, Austria.

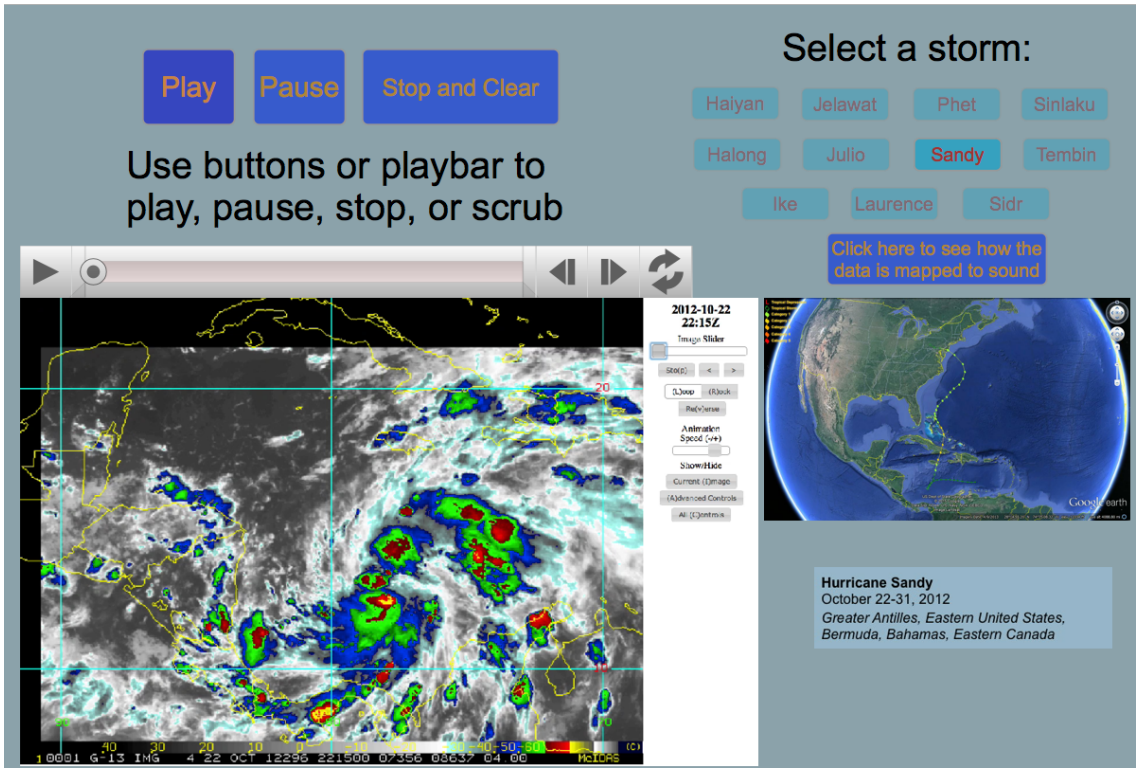


Figure 1: Interface for storm videos and sonifications exhibit (created in Max/MSP)

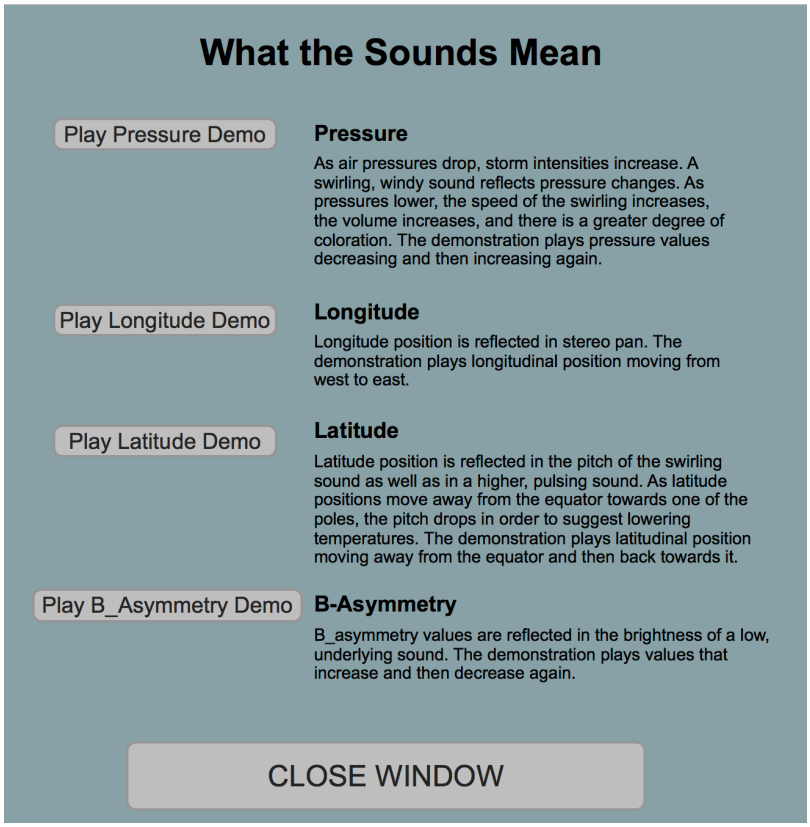


Figure 2: Tutorial screen from storm sonification kiosk

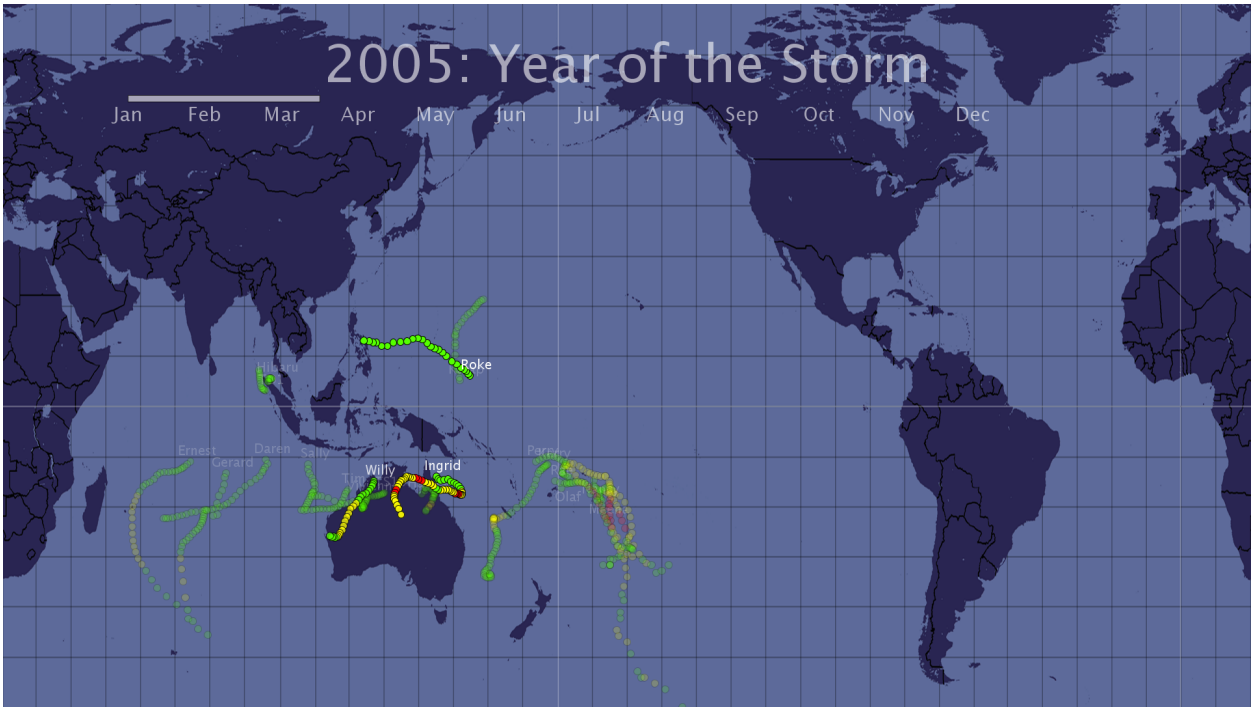


Figure 3: Example frame from Spaghetti of Storms video

SONICALLY-ENHANCED SMART EXERCISE APPLICATION USING A WEARABLE SENSOR

Eric Vasey[1], Byungjoo Noh[2], Tejin Yoon[2], Myounghoon Jeon[1]

Mind Music Machine Lab,
Michigan Technological University,
[1]Department of Computer Science, [2]Department of Kinesiology and Integrative Physiology,
1400 Townsend Dr., Houghton, MI 49931 USA
{emvasey, bnoh, tyoon, mjeon}@mtu.edu

ABSTRACT

The Smart Exercise application is an Android application paired with a wearable Bluetooth IMU sensor that is designed to provide real-time auditory and visual feedback on a user's body motion, while simultaneously collecting kinematic data on their performance. The application aims to help users improve physical and cognitive function, improve users' motivation to exercise, and to give researchers and physical therapists access to accurate data on their participants' or patients' performance and completion of their exercises without the need for expensive additional hardware.

1. INTRODUCTION

The application has two main functions. First, the Lunge Piano (Figure 1) gives real-time audiovisual feedback on performance in a clock-lunge exercise. Users rotate their body to face a direction to select a note, and perform a correct lunge motion to play the note they selected. The Lunge Sequences functionality integrates a memory game into the piano. Users need to remember the given sequence and use the Lunge Piano to produce notes in a particular sequence from memory. For example, if the user is given the note sequence "CDFE", they must rotate their body to select each note in the sequence in order, and successfully perform a lunge on each of those notes. Data on the user's attempts to complete sequences, such as times, number of attempts, correct attempts, and cumulative errors are logged to the Android device for later viewing. The second function is the Squat Muter, which has users balance on one leg while holding a squatting posture for as long and as steady as possible. As the user starts to lose balance, the app will warn them using periodic beeps. If the user is unstable for longer than a specified time interval or loses their balance completely, the app will give a series of rapid higher pitched failure beeps, indicating a failure condition.

2. RELATED WORKS

Rosati et al. investigated the effect of audio feedback on a motion tracking exercise [1]. Participants moved a pen along a touchpad and 2 degree of freedom joystick, to follow a dot on the screen while receiving differing combinations of task-related, error-related, and visual-related feedback. They found that auditory feedback for upper-body exercises was beneficial, task-related feedback improved performance more than error-related feedback, and auditory feedback on the motion of the target is better than auditory feedback on controller motion.

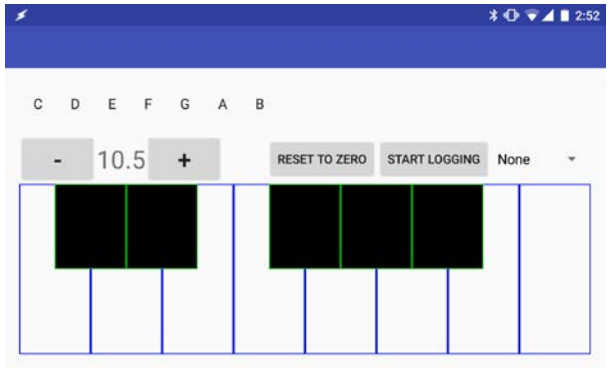


Figure 1: Smart Exercise Application

3. TECHNICAL DETAILS

The Android application collects data from a Metawear C/CPRO wearable Bluetooth IMU sensor from *Mbientlab*. When data is received, the phone passes the data through a threshold filter allowing through data that is > ±5% of calibrated values, followed by a Kalman filter to remove noise in the data.

4. CURRENT AND FUTURE WORK

A validation study (not yet published) has been performed showing high correlations between the X ($r = 0.91 \pm 0.03$, $p < 0.001$), Y ($r = 0.91 \pm 0.04$, $p < 0.001$), and Z ($r = 0.92 \pm 0.02$, $p < 0.001$) axis accelerometer data compared to a high-grade commercial accelerometer. The direction angle for note selection in the clock lunge exercise also showed high correlation ($r = 0.99 \pm 0.01$, $p < 0.001$) with a camera based motion analysis system. Another usability study has also been conducted, with approximately 90% of participants reacting positively to the application. Our future work involves evaluating the audio and visual feedback of the application to investigate its usefulness and intuitiveness.

5. INSTALLATION REQUIREMENTS

This installation would consist of the Smart Exercise application running on a smartphone with the wearable

Bluetooth IMU. A headset would be used as well to allow for higher clarity of the tones. In order to minimize the risk of injury, we would need a space of at least 5 feet x 5 feet for audience members to try the application. Audience members will be able to clip the wearable sensor onto their clothing, connect to the sensor using our application, and use the Lunge Piano and Squat Muter functions.

Here's the link to a demo video: https://drive.google.com/file/d/0B5q7ZwXykkh6Y2ZZS2ZVUkNxVFk/view?usp=sharing

6. REFERENCES

[1] G. Rosati, F. Oscari, S. Spagnol, F. Avanzini, and S. Masiero, "Effect of task-related continuous auditory feedback during learning of tracking motion exercises," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, no. 1, pp. 79, 2012//, 2012.

#416: REAL-TIME TWEET SONIFICATION FOR REMEMBRANCE OF THE SEWOL FERRY

Ridwan Khan[1], Myounghoon Jeon[1,2]

Mind Music Machine Lab
Michigan Technological University,
[1]Department of Computer Science, [2]Department of Cognitive and Learning Sciences,
1400 Townsend Dr. Houghton, MI 49931 USA
{ridwank, mjeon}@mtu.edu

1. INTRODUCTION

1.1. Goal

There was a tragic accident of sinking a ferry in Korea three years ago. We aim to remember this tragedy together to support people who lost their family in the accident and let more people know about it for prompt resolution of this matter. Korean artists and scholars have tried to show their support for family by performances and writings. In the same line, we have created a real-time tweet sonification program using "#416", the date of the tragedy. The text of tweets including #416 is translated into the Morse code and sonified in real-time.

1.2. Background

The South Korean ferry, "Sewol" sank on its voyage from Incheon to Jeju Island on April 16, 2014. The accident claimed the lives of more than 300 passengers. Of the 476 people aboard, 325 were students from Danwon High School and 250 students and 11 teachers passed away in this tragic accident [1]. There are still nine missing people. One of the most miserable parts was that passengers had been told to remain in their seats and may have stayed there until it was too late.

The tragedy, among South Korea's worst peacetime disasters, led to criminal convictions, the resignation of the country's prime minister and the death of the billionaire religious leader who owned the ferry [2]. However, the cause of the accident (and more importantly, why they were not saved in time) is still not clear, whereas people are suspicious of collusive ties among business and governmental authorities.

1.3. Motivation

Our sonification would be people's voice, "We remember this tragedy and will never forget"; it would be victims' family and parents' voice, "We are desperately waiting through the night"; when it is translated into the Morse code, it would be victims' voice for help, "We are still here, under the water. Please come and save us"; and it would be all of our voice, "We will not just remain doing nothing, but stand against injustice."

1.4. Related Works

As research on data sonification [3] has increased, data from social network services have been used for real-time sonification. For example, Dahl, Herrera, and Wilkerson [4] have created "TweetDreams", which is a real-time sonification and visualization of tweets. In their program, tweets including specified search terms are associated with each other and made into networks. Then, the networks make a new melody using concepts, such as inheritance and mutations from the parent melody. Hermann and his colleagues [5] introduced tweetscapes at ICAD 2012. Tweetscapes provides characteristics of Twitter messages, including density, origin, impact, and how topics change over time. Despite these earlier works, little has been done on use of real-time sonification of social network services data for promoting awareness of social issues in the auditory display community, including ICAD.

2. TECHNICAL DETAILS

2.1. Technical Details

Our program is an application to fetch tweets from Twitter and make sound out of these tweets. We have used Twython (https://twython.readthedocs.io/en/latest/), a python package to fetch tweets, which essentially uses twitter search API (Application Program Interface) (https://dev.twitter.com/rest/public/search). Given a keyword (e.g., a hashtag: staring with "#"), the application fetches tweets of that keyword. We have also used another python package, Apscheduler (https://apscheduler.readthedocs.io/en/latest/) to schedule the fetching of tweets in an interval. Based on time (30 seconds in our case), we get the latest tweets for that keyword. Thus, whenever people tweet using the keyword, it will be fetched by our application in a few seconds.

2.2. Tweet Sonification Version 1.0

To sonify the text of the tweet, including the keyword (in our case, "#416"), we used another python package, MidiUtil (https://pypi.python.org/pypi/MIDIUtil/1.1.1). We have customized sound parameters, including pitch and channels,

and then generated a MIDI (Musical Instrument Digital Interface) sound file which holds musical notes. Each character of the text of the tweet is translated into a number, which represents the pitch for the note. Also, the position of the character in the tweet text is mapped to a channel (Figure 1). In this way, the text is sonified into a collection of notes. These musical notes are written into a MIDI file and played.

```
#Import the library
from midiutil.MidiFile import MIDIFile
import os

class MidiConverter:

    def stringToMidiConverter(self, musicString):
        #musicString = 'Mind Music Machine Lab'
        countChars = len(musicString)

        print "The lenght of the musicString is %d" % (countChars)

        musicMIDI = MIDIFile(countChars)
        duration = 1
        volume = 100
        track = 0
        time = 0


        for ch in musicString:
            musicMIDI.addTempo(track,time, 120)
            channel = (time%15)
            pitch = ord(ch)
            musicMIDI.addNote(track,channel,pitch,time,duration,volume)
            track += 1
            time +=1

        midiFilename = "tweet.mid"
        binfile = open(midiFilename, 'wb')
        musicMIDI.writeFile(binfile)
        binfile.close()

        os.system('timidity '+ midiFilename)
```

Figure 1: Python code for real-time tweet sonification.

2.3. Tweet Sonification Version 2.0

After the first design iteration, we have pondered about aesthetics of the application and how to include multiple meanings of our sonification. Then, we have developed another version to better reflect diverse voices (specifically, voice of victims who are still waiting for the salvage of the wrecked ferry under the water) as mentioned in Motivation. In this version, instead of mapping characters of the tweet text to pitch and positions of characters to sound channels, we have transformed the text to a Morse code string. The Morse code represents each letter of the alphabet using short and long sounds, dots and dashes. Based on this, each character is transformed into a collection of dots and dashes. These collections are concatenated to form of the Morse code for the tweet text. Then, we play a sound composed of short (dot) and long (dash) parts.

3. CURRENT WORK

We are currently working on the development of a graphical user interface of the application. The basic idea for future improvement is to show the tweet and encoded string, and visualize the sonified sound on the screen. We expect we can demonstrate this completed version at the conference.

4. INSTALLATION AND FUTURE PLAN

This installation would consist of our laptop to visualize and sonify the tweets. We will use our headsets for sonification. Audience will use their own smartphone or our laptop to try to tweet text using #416 and experience our application in real-time. For this installation, we will only need a small table for our laptop. After the installation at ICAD, we will demonstrate this installation at the Michigan Tech Demo Space.

Here is the link to our demo video: https://drive.google.com/open?id=0B-x6yGv_leBwZHNyWHVRZ3B0S28

5. REFERENCES

[1]https://ko.wikipedia.org/wiki/%EC%84%B8%EC%9B%94%ED%98%B8_%EC%B9%A8%EB%AA%B0_%EC%82%AC%EA%B3%A0

[2]https://www.nytimes.com/interactive/2015/04/12/world/asia/12ferry-timeline.html

[3] Hermann, T. (2008). Taxonomy and definitions for sonification and auditory display. In Proceedings of the 14th International Conference on Auditory Display (ICAD 2008).

[4] Dahl, L., Herrera, J., & Wilkerson, C. (2011). TweetDreams: Making music with the audience and the world using real-time Twitter data. Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 272-275. 30 May - 1 June 2011, Oslo, Norway.

[5] Hermann, T., Nehls, A. V., Eitel, F., Barri, t., Gammel, M. (2012). Tweetscapes: Real-time sonification of Twitter data streams for radio broadcasting. Proceedings of the 18th International Conference on Auditory Display, pp. 113-120. Atlanta, GA, USA, June 18-21, 2012.

Antonio D'Amato *Une rencontre*

τὸ ἀντίξουν συμφέρον καὶ ἐκ τῶν διαφερόντων καλλίστην ἁρμονίαν καὶ πάντα κατ' ἔριν γίνεσθαι
What opposes unites, and the finest attunement stems from things bearing in opposite directions, and all things come about by strife.

μεταβάλλον ἀναπαύεται
It rests by changing.

ξυνὸν γὰρ ἀρχὴ καὶ πέρας ἐπὶ κύκλου περιφερείας
Concerning the circumference of a circle the beginning and the end are common.

Herakleitos of Ephesos Fragments 8, 84 and 103

Une rencontre (2013) is intended as a "virtuoso" piece where short vocal samples are deeply processed together with acoustic instruments samples and synthesized elements in order to achieve or at least tending toward a total fusion of these elements. It is not just a mixing work, but instead it is the fusion of more sound objects in one, mainly through convolution algorithms. In this way I especially researched for a spectral merging of voice and acoustic or electronic instruments.

The goal is to attempt forcing the merging of communicative strengths coming from not homogeneous sources into a single flux.

I intentionally propose an artificial procedure where the allocation of formantic regions is extracted from audio samples of singing voices, and subsequently superimposed to audio samples of traditional acoustic instruments or even to synthesized sounds. The original formants of the acoustic and synthesized sounds have been partly preserved in some moments, while the formants coming out from voices have been added or sometimes have replaced them, therefore in few points an overload of harmonics has been deliberately caused in order to experiment an emphasis of the expressive strength and emotional communication of the resulting timbre. The work involves also the use of other DSP processes throughout: selective spectrum saturation, resonant notch bank filtering, multiple delays, and distortion. In some isolated moments the output of a process is injected in a granularization module, in order to achieve a subtle and mutant texture.

The title reveals the aesthetic sense of the work, where a relevant meeting can be not just an encounter, but also the occasion for a mutual exchange, therefore everything and the whole of the world are in continuous transformation, by means of mutual confrontation. In this way a symmetric duality established in any action involving a mutation: the acting subject that changes an object is reciprocally transformed by the object itself.

CONCERT

Antonio D'Amato
Körper

abstract

Körper is an acousmatic piece entirely based on the elaboration of an acoustic pulse sequence which was produced in the course of a MRI diagnostic test. The aesthetic idea implied in the composition refers to the topical and controversial theme known as "global control and censorship". Through the examination of the constant and continuous information flow, which is either consciously or unconsciously produced by everyone, it is possible to accomplish a condition of control; that condition ought to benefit the national and global security.
The question is: How deep or intrusive should the control on individuals be in order guarantee global well-being and security?
At the moment, there is no univocal answer. Control is not only coercive - it can take the pleasant and subtle appearance of a custom-tailored set of information, directly injected into the communication flows usually employed by individuals. As the occasional intrusion into the personal communication flow is barely noticeable, could the large mass of information, imposed to the users of digital devices, produce a sort of control on the behavior and habits of the individuals? Has that sort of control only a commercial intention?
If nowadays cameras and sensors constantly watch movements of the individuals in cities and buildings, can we assume that in the future cells and chemical reactions in our bodies will be scanned and examined in order to gather information to be collected, stored and processed?

Technically speaking the composition uses exclusively a short audio recording of a MRI test.
A large number of processes and signal elaboration modules are applied in order to subdue the crude audio sample to the compositional requirements.
Spectral editing and resonant filters are chained in order to isolate restricted areas in the whole sound object. The foundation material is clearly revealed only at the very end of the piece. The piece was composed at ZKM studios in Karlsruhe.

Composer: Stephen Roddy
Piece: The Good Ship Hibernia
Time: 5:54
Type: Fixed Media Stereo Soundscape Sonification

Description: "The Good Ship Hibernia" is an *Embodied Soundscape Sonification*. It explor es the impact of the world-wide financial crash on one of the harder hit nations in Eur ope, Ireland. It uses both soundscape and harmonic materials and employs sonification techniques to reflect the World Bank's figures for Irish GDP growth rate from 1979 to 2013. The piece is structured around a number of embodied metaphors (as defined by Lakoff and Johnson 1980) and an embodied balance schema (as defined by Johnson 1987). It uses the metaphor of maritime journey where "smooth sailing" and "good weather represent "good times" and "rough seas" and "bad weather" represent "bad times". While the growth rate is strong the sailing is smooth and weather is good. When the growth rate shrinks the weather becomes stormy and seas become rough. The sense of balance in the soundscape shifts in accordance with the data. Harmonic material also sounds thr oughout the journey. This material was performed in response to the soundscape of the sea journey with the perceived fidelity and timbral character of the performance determined by theGDP data.

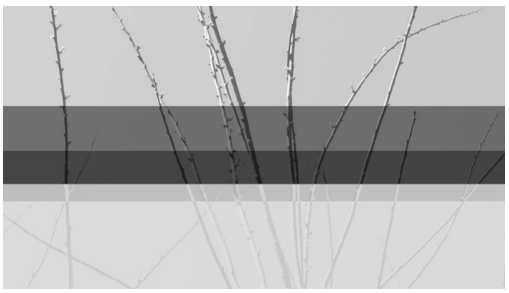
Technical Requirements & Stage Layout: This is a fixed media piece composed for two-channel stereo diffusion. The piece requirements for the piece are minimal. It can be di ffused using any even number of paired speakers as long as the left right orientations are maintained and the speakers produce good sound quality (preferably a pair of G enelec 404A's). The piece can be supplied by the artist on numerous media such as USB keys, CD/DVD's or external hard drives or it can be diffused directly by the artist thr ough the sound desk via external sound card and laptop. The piece can be supplied in .wav or .aif formats.

Instrumentation & Number of Performer: No performer/instruments required

Alfredo Ardia

Rami

sound, video
Italy, 2015



The idea behind *Rami* came during a study about beats phenomenon: an acoustic interference, produced by two or more sine waves with slightly different frequency, which results in a perdiodic amplitude modulation.
Looking at a naked tree I imagined it as a score where a set of frequency moving near and away, crossing each others, following the branch's shape and creating complex beats patterns.
In *Rami* the sound is created by a sonification process of shape and movements of the tree showed in the video component.

Andrew Litts - *Singularity for trumpet and electronics*

Singularity intertwines the individual with its role in society. Through live recording and manipulated playback, the solo trumpet is heard with alternate versions of itself, as if the digital medium distorts, misconstrues, and turns the original voice on its head, much as is done on social media. Certain portions of the trumpet part are recorded into buffers in the MAX patch and the computer plays back myriad versions of these recorded segments in transposition, with time stretching, aleatorically, and micro-polyphonically to alter the original voice and propose commentary on the sense of belongingness digital space allows for an individual.

Digital media allows for the dissemination of ideas and conversation at an ever-increasing rate and ease. *Singularity* joins in dialogue with other "augmented instrumental" pieces in that the computer acts both as an extended voice of the original instrument and an allegory for society that conflates the role of the individual (the layering of trumpet voices even mimics the writings of Gabrieli and his use of antiphonal brass, pieces of which serve as divine homages from earthly beings conveying their messages collectively). The vehicles to those who long lacked a voice are an undisputed upside to the modern propagation of technology. The downside, though, is the potential danger in the loss of identity as the echo chamber of the Internet turns into a place for hyper-individualism to contribute to an amalgamation of noise. As we listen to each other less, ideas mean less.

Julius Bucsis - *Portraits of Nine Revolving Celestial Spheres (audio)*

Portraits of Nine Revolving Celestial Spheres was inspired by Gustav Holst's composition *The Planets*. However, unlike *The Planets*, in which Holst intended to convey concepts related to astrology, *Portraits of Nine Revolving Celestial Spheres* utilizes the sonification of astronomical data derived from scientific observations as the basis for the musical elements. Each section of the piece represents one of the planets of the solar system and is built from a set of data corresponding to each planet respectively. Among the included data are measurements related to albedo, atmosphere, distance, diameter, gravitation, magnetic field, orbital factors, pressure, relative abundance of the most common elements, and temperature. There are nine sections because as of this writing, there is considerable evidence based on orbital anomalies suggesting the existence of an undiscovered planet. Since there is no data about this hypothetical planet, its corresponding section in the piece employs a nebulous process in the selection of musical materials in order to suggest a broad range of possibilities. The title pays homage to Nicolaus Copernicus' seminal work, *On the Revolution of the Heavenly Spheres,* which challenged the widely held view of the time that the earth was at the center of the universe.

After Images

(2017)

By Roberto Zanata

“After Images” is an audio/video work generated by a given pattern using various node data. An **afterimage** is a non-specific term that refers to an image continuing to appear in one's vision after the exposure to the original image has ceased. It hase been realized with a patch in Max/Msp that it allows to use jitters visual effects for high quality 2d images.

Short Bio:

Roberto Zanata born in Cagliari, Italy where he also graduated in Philosophy. A composer, musician and musicologist in electronic music, he studied and graduated in composition and electronic music at the Conservatory of Cagliari. In the middle of nineties Roberto became active in Italy and abroad. He wrote chamber music, music for theatre, computer music, electroacoustic and acousmatic music as well as multimedia works. His music is published by Audiomat, Taukay and Vacuamoenia. In International competitions his works have been awarded *Grands Prix Internationaux de Musique Electroacoustique* (Bourges), *Interference Festival* (Poland), *Sonom Festival* (Mexico) and more. He actually teaches Electronic Music at the Conservatory of Foggia (Italy).

Roberto Zanata
Via Satta n. 68
09127, Cagliari
Italy

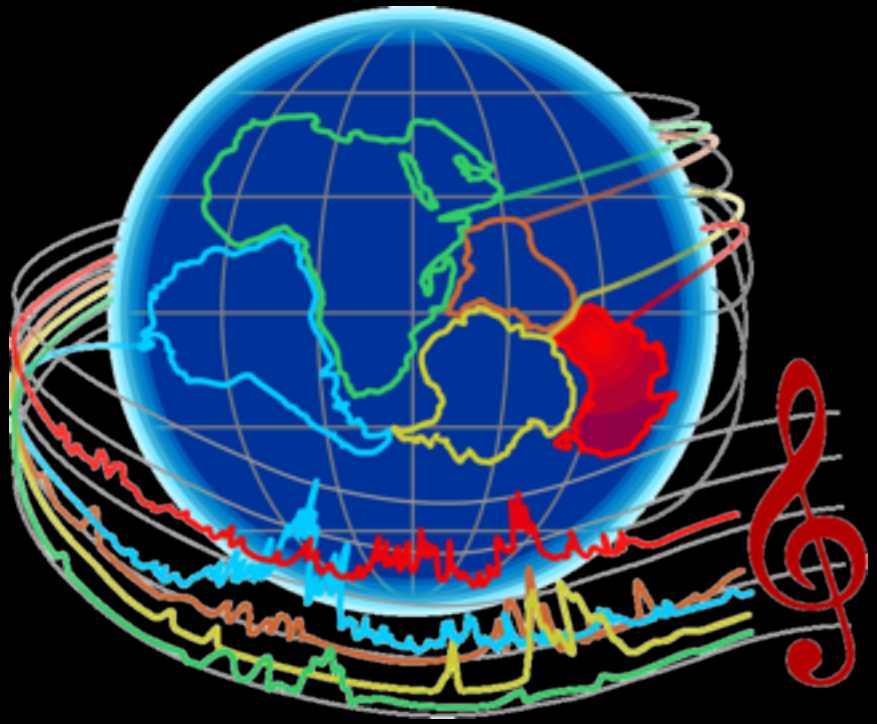
Email: robertozanata@gmail.com

Eonsounds: Fiamignano Gorge

A Tale of Two Mountains

James Cave, AHRC-funded PhD candidate in Music, University of York

EONSOUNDS.ORG



INTERDISCIPLINARY COLLECTIVE:

James Cave, Tim Ivanic, Jude Brereton

Ben Eyes, Supriya Nagarajan, Fawna Korhonen, Hugo Janacek, Daniel Quinn, (plus others)

University of York

cmrc YORK

The University of Edinburgh

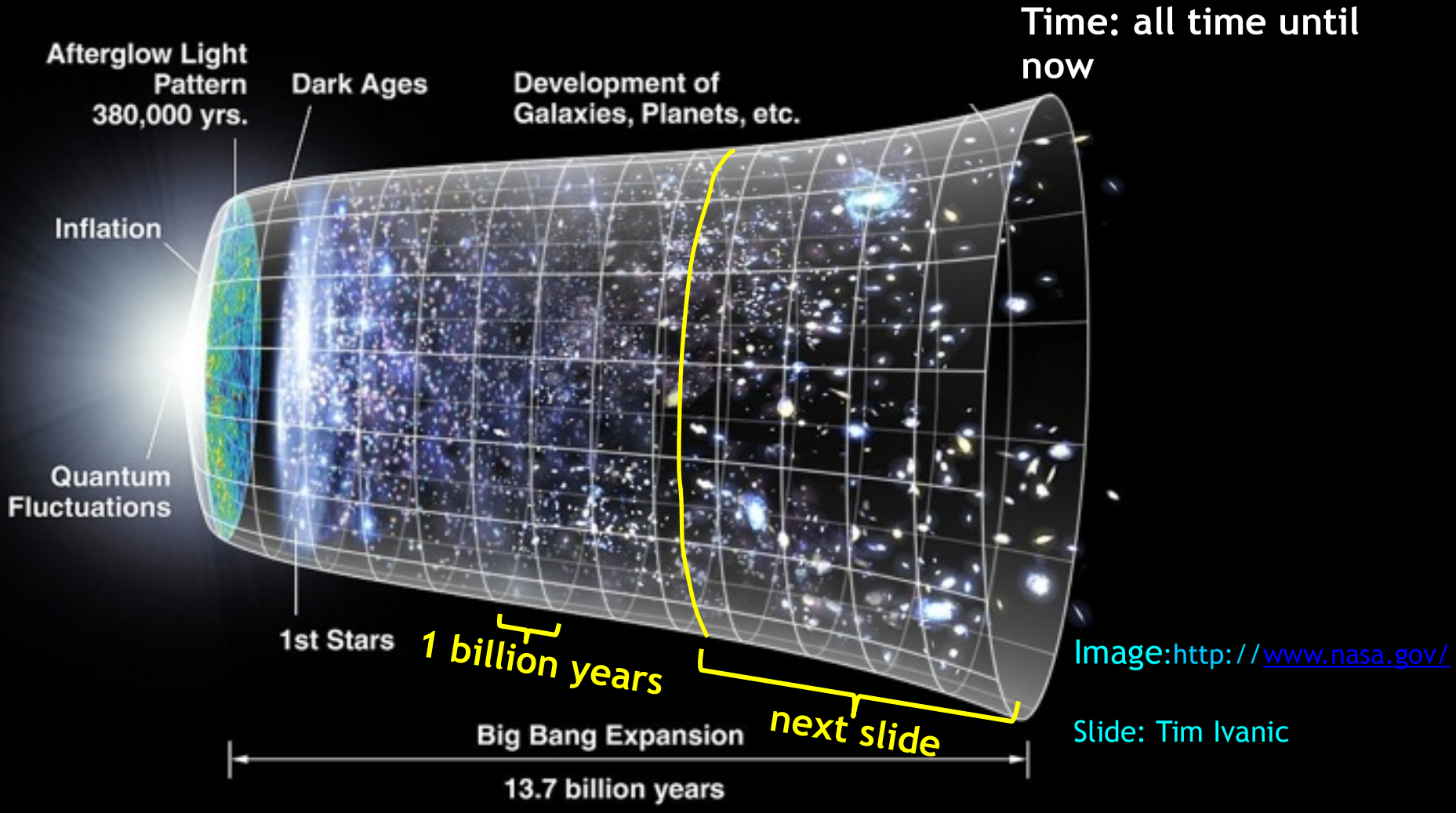
York Festival of Ideas

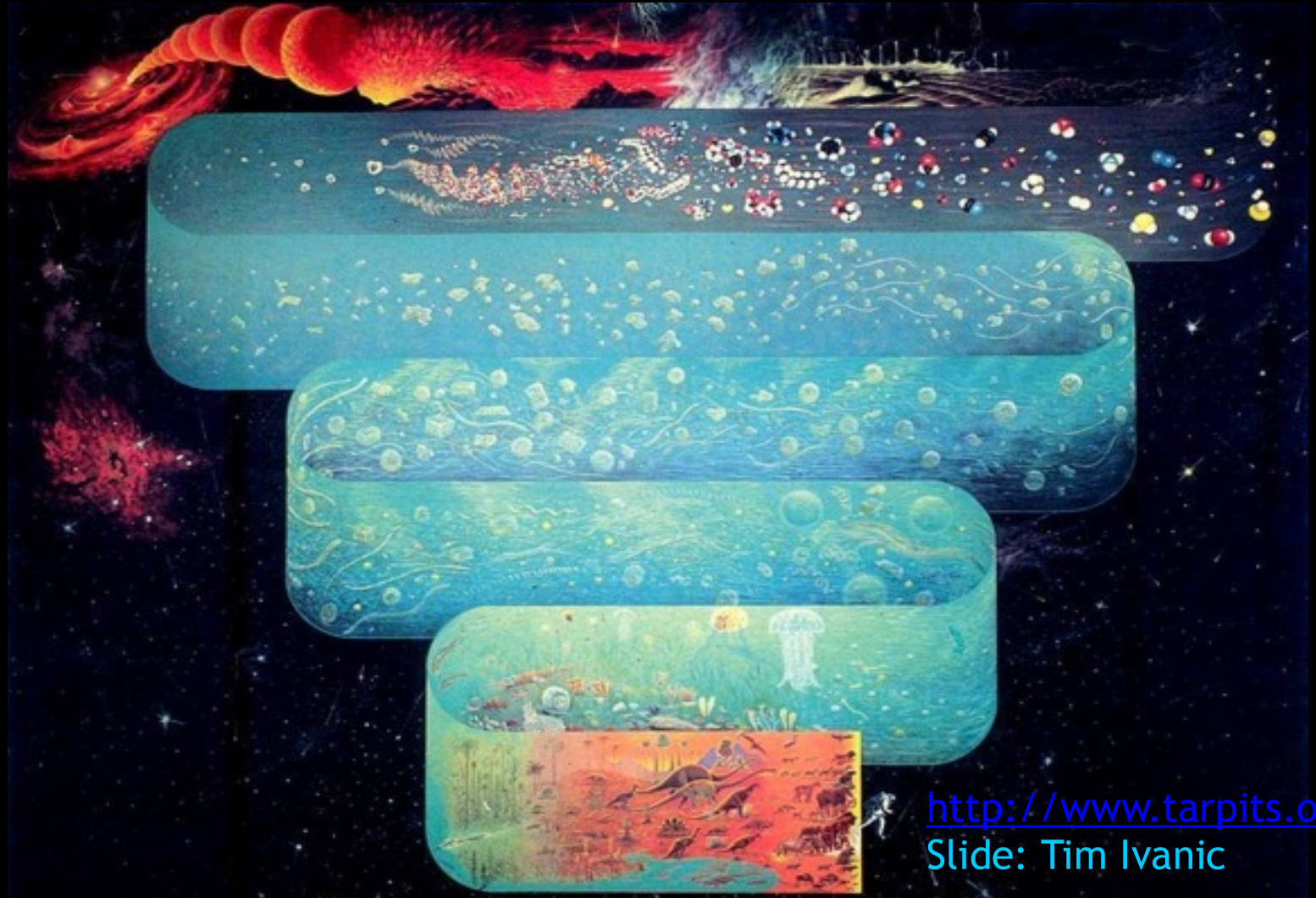
manasamitra
introducing south asian arts to generations

It all started here, with the question of 'deep time'

(Image: Banff Centre)

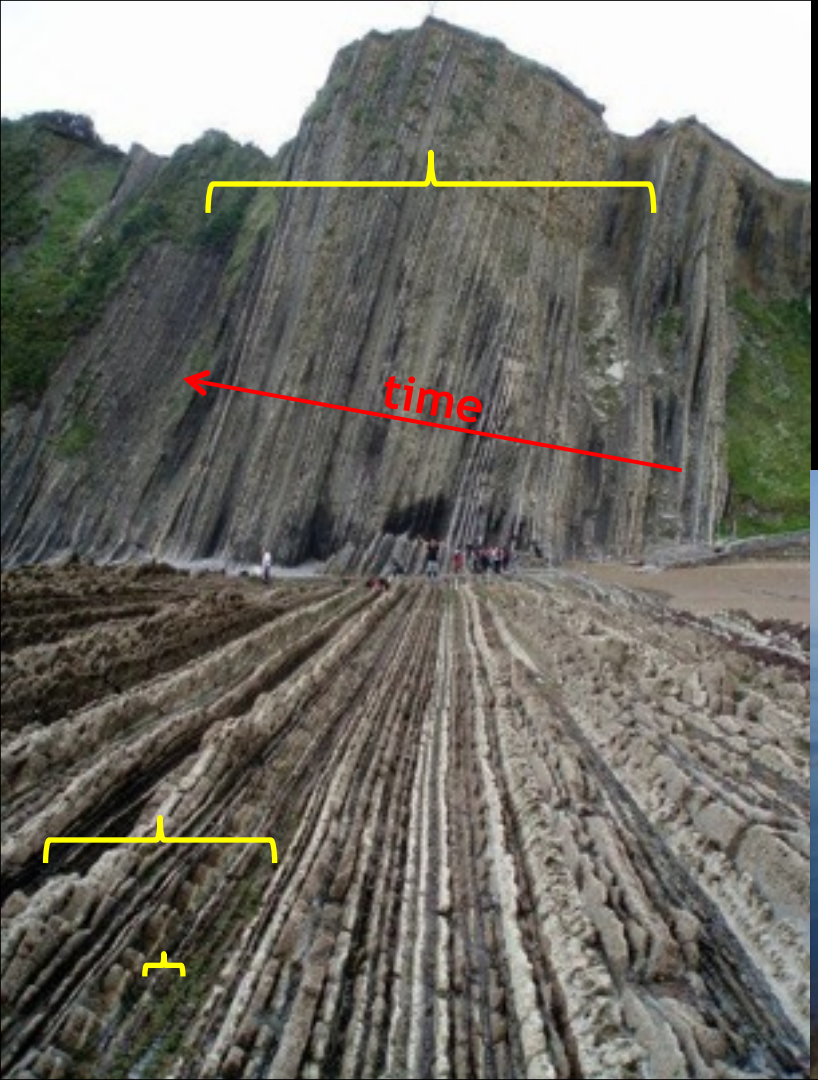
Early attempts at music-making…


Time: all time until now
Afterglow Light Pattern 380,000 yrs.
Dark Ages
Development of Galaxies, Planets, etc.
Inflation
Quantum Fluctuations
1st Stars
1 billion years
next slide
Big Bang Expansion
13.7 billion years
Image:http://www.nasa.gov/
Slide: Tim Ivanic



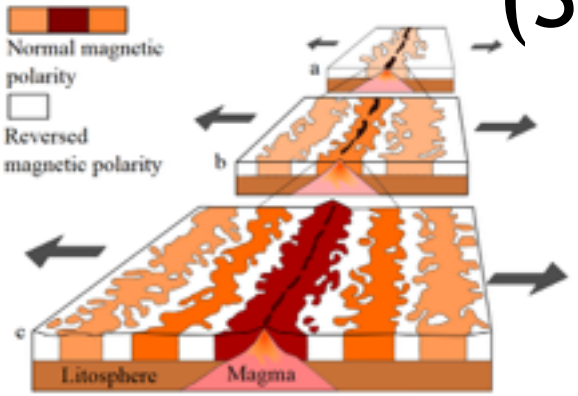
http://www.tarpits.org/
Slide: Tim Ivanic

A variety of geological ‘scores’

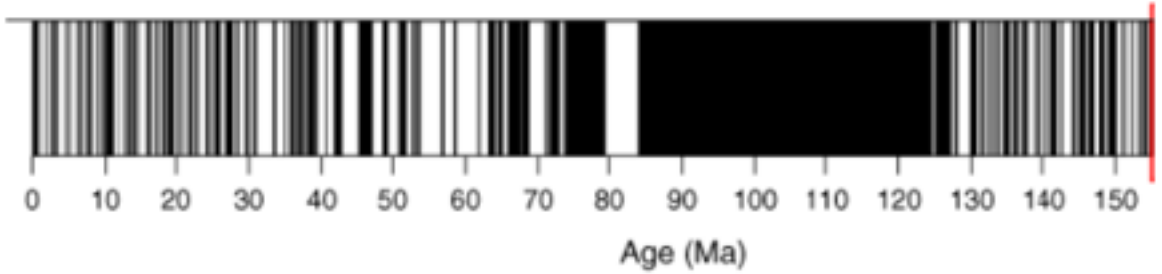


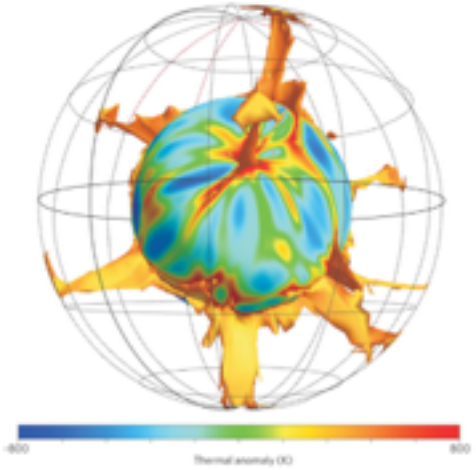
Reading rocks
on all scales
=relative time
Slide: Tim Ivanic

‘Recordings’ of the globe (Slide: Tim Ivanic)

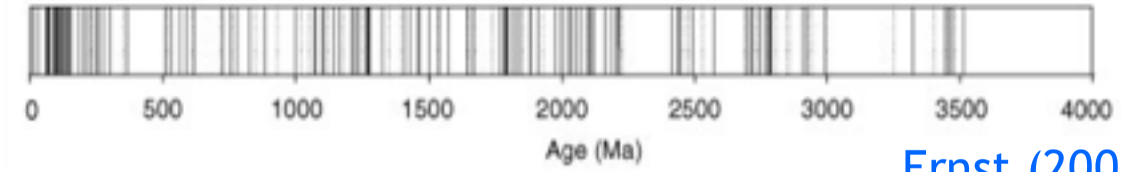


Geomagnetism





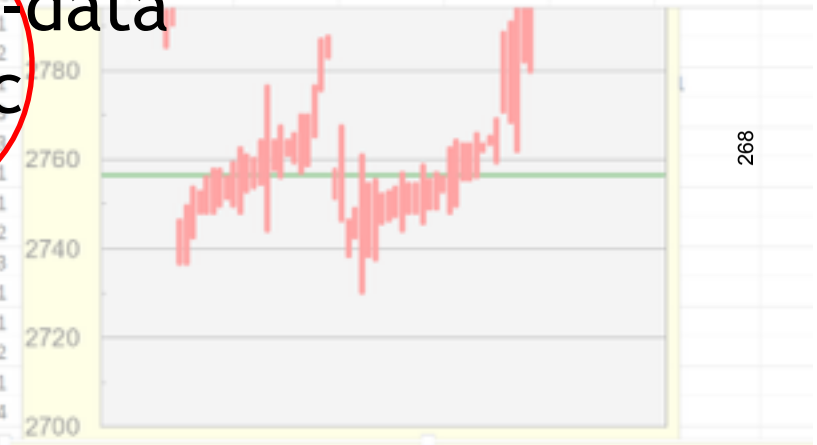
Deep mantle plumes

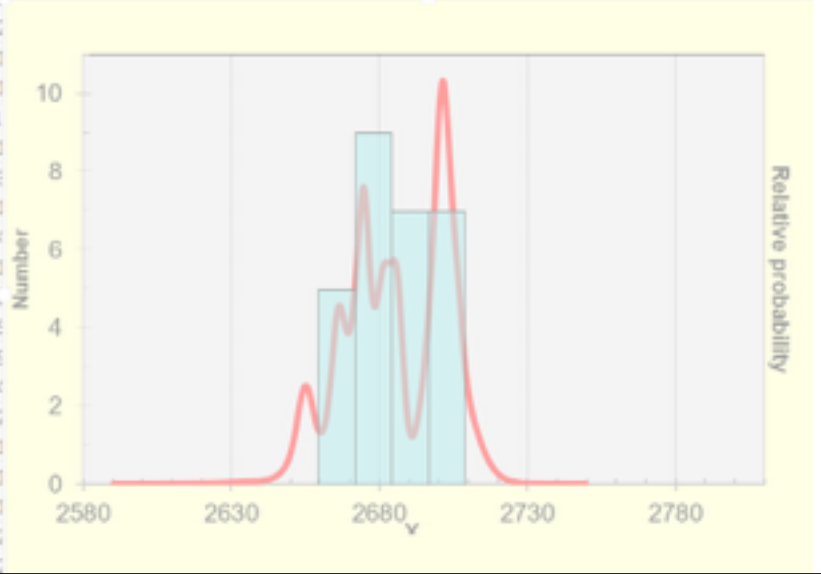


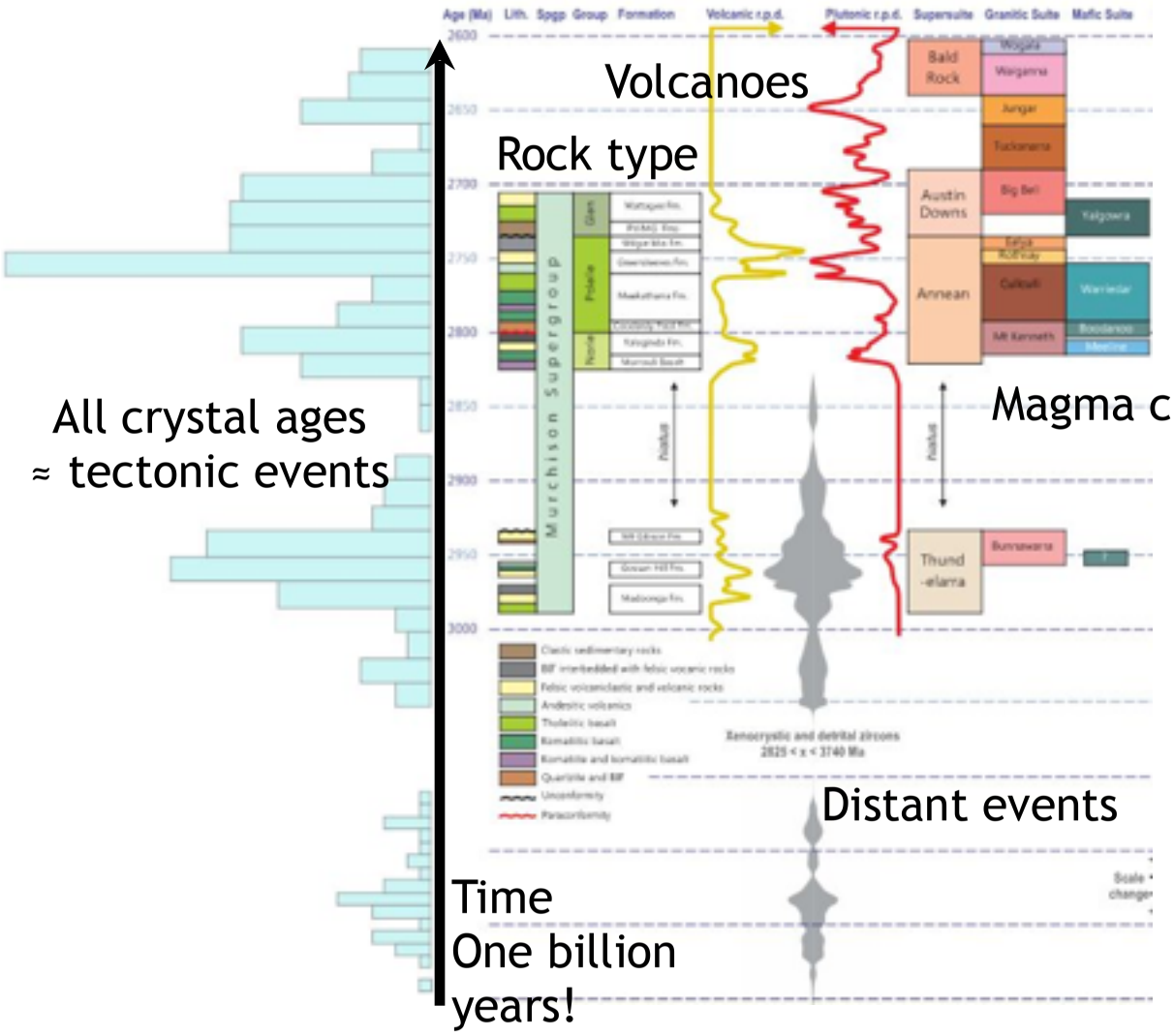
Ernst (2003)

Geological field-data
Slide: Tim Ivanic

SUITE	Locality	Sample #	Age			Lithology		
METAGRANITE	Wogala	185928						
METAGRANITE	Saturn Pit, Mount Ma	97124	2735	8	felsic porphyry	577314	6901120	2
METAGRANITE	Weld Range	185926						
METAGRANITE	Wurarga	99969164						
METAGRANITE	Noongal metagranite	99964100	2741	4	metamonzogranite	474807	6886252	3
METAGRANITE	Coodardy	178194	2746	3	metagranite gneiss	566032	6987547	1
METAGRANITE	N Reedy	178141		5	qtzdiorite	628441	7001698	1
METAGRANITE	Peter Well gdr	93991	2747	6	aplite	615536	6974399	2
METAGRANITE	Wurarga S/Cagacarrc	99969142	2747	3	metagranite	436579	6830649	3
METAGRANITE	Mt Mulchahy	185920	2748	5	metagranite	572484	7005055	1
METAGRANITE	Glen	178196	2748	3	granodiorite	558474	7000166	1
METAGRANITE	N Peter Well gdr	93990	2749	4	granitoid	615570	6978092	2
METAGRANITE	BADJA	155822	2749	3	granite gneiss	485369	6835504	1
METAGRANITE	Meekatharra	06GNH02	2750	4	Fld-Hbl porphyry	641300	7031500	4
METAGRANITE	Boomer Pit	92283	2752	3	porphyry	580236	6897043	1
METAGRANITE	N of Peter Well gdr	93989	2752	4	granitoid	615588	6979938	2
METAGRANITE	Mt Mulgine	207628	2752	3	metagranite	497923	6771884	1
METAGRANITE	BADJA	155858	2752	13	granite	481361	6815692	1
METAGRANITE	S Tallering	89457	2753	10	granite	365080	6884900	1
METAGRANITE	Mercator	185958	2755	6	microgranite	645310	7037839	1
METAGRANITE	Mt Mulgine	99964016	2756	20	granite	497943	6771112	3
METAGRANITE	Coodardy	185921	2757	4	monzogranite	558949	6996423	1
METAGRANITE	Meekatharra	06GNH01	2758	7	Qtz-Fld porphyry	641300	7032000	4
METAGRANITE	Peter Well gdr	185933	2759	2	granodiorite	615014	6974387	1
METAGRANITE	W Gabanintha	96129	2759	7	internal granite	660531	7021984	2
METAGRANITE	Cue - Queen of May	W378	2759	4	tonalite	590757	6968914	5
METAGRANITE	Norie Pluton	PFG1	2760	8	tonalite	634309	7029666	5
METAGRANITE	Meekatharra	06GNH03	2761	7	Mafic granite	642300	7032500	4
METAGRANITE	Gabanintha	96129	2769	7	granite	660673	7022120	2
METAGRANITE	Nannine granodiorite	178104	2782	7	granodiorite	628483	7035709	1
METAGRANITE	Meekatharra	178102	2781	3	metatonalite	629667	7041124	1
SEDIMENTARY	Weld Range South	185925	2740	4	quartzite	589333	7020101	1
SEDIMENTARY	Golconda	96121	2780	14	chert	633112	7035340	2
SEDIMENTARY	Gabanintha	93974	2788	30	shale	655955	7042280	2

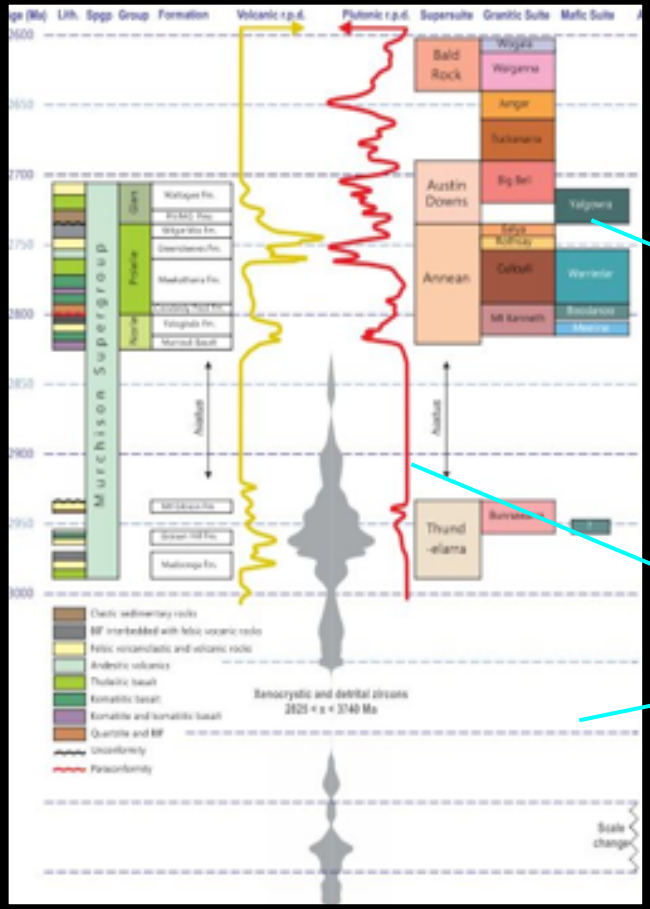


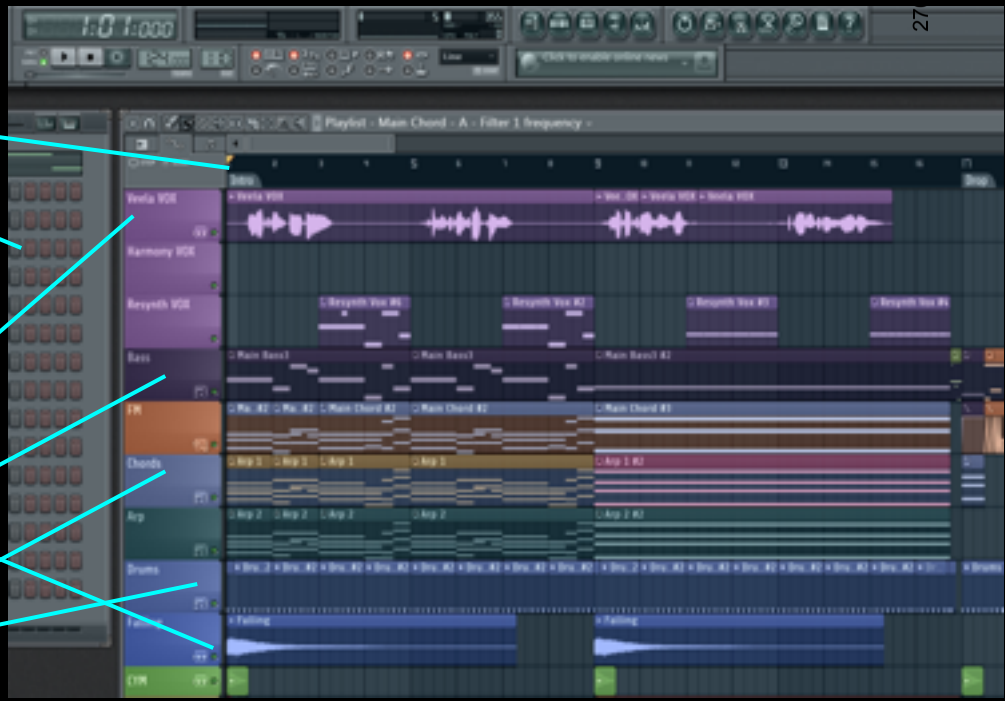




A portion of
Western Australian
Geology from my mapping
area 500 km north of Perth

Geological time → Musical time





Compose music from geology

mountains, oceans, volcanoes, earthquakes

Why sonify?

- Arguably easier to present events in space and time at the same time
- Possibility of using sound-type and density to represent rock-types and formations
- More accessible to certain audiences (e.g. people with visual impairments)
- Opportunity to use University of York research expertise in ambisonics and sonification
- Cross-pollination between artistic and scientific applications of sonification in order to find new directions for research and public engagement
- Explore links to particular musical styles and traditions such as Indian classical ‘tala’

What is sonification?

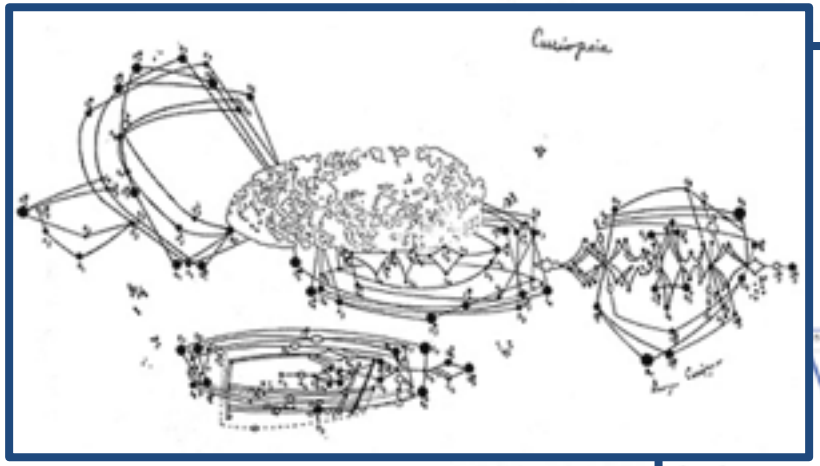
- Let’s try some…
- Apples, oranges, pears
- Red apples, blood oranges, prickly pears

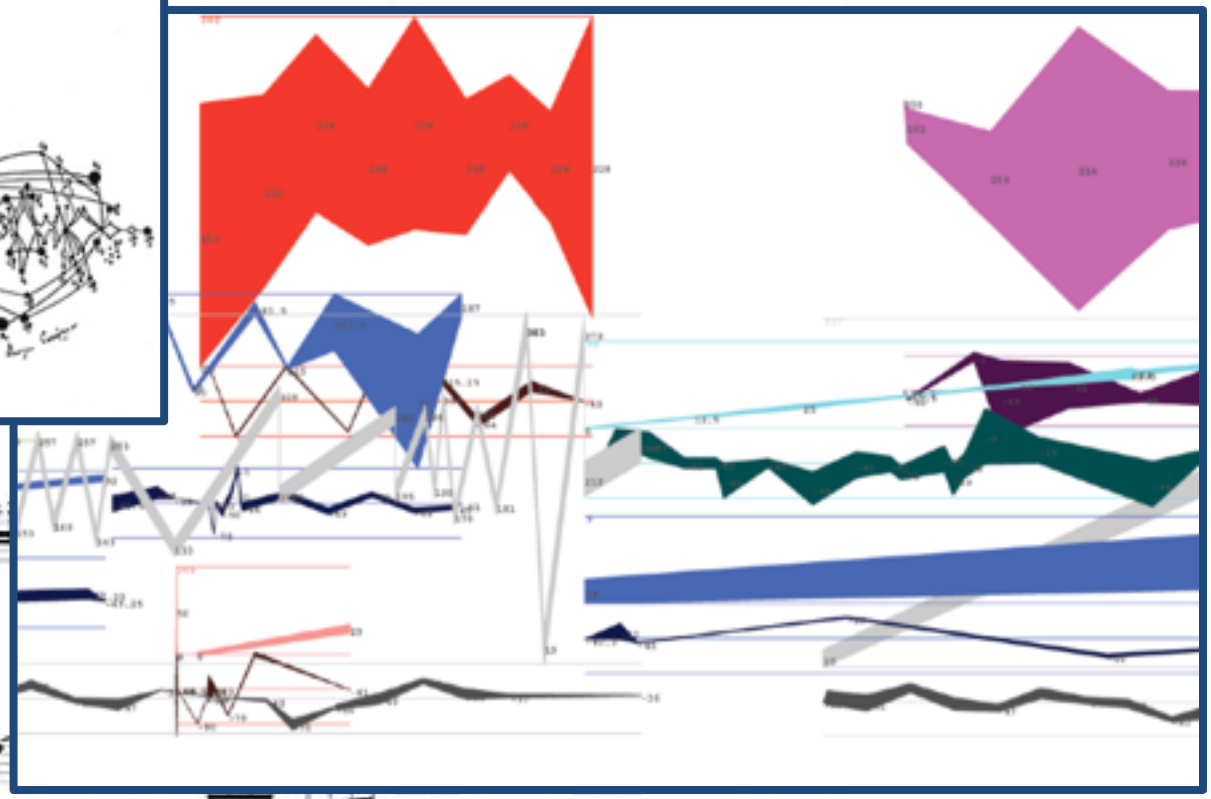
Different approaches from our initial development day (June 2016)

- Simple drone: (Daniel Patrick Quinn) created by mapping broad "event intensity" mapped to volume. Used as backing track for more freeform improvisations
- Multi-channel drone (Duncan Chapman)
- Pure sonification (Ben Eyes et al): transformation of single geological datasets using MAX/MSP
- Graphic notation: (James Cave, Ann Warde, Jacob Thompson-Bell). Use of crystal images and geological event timelines to create graphic scores

‘It sounds like Depeche Mode’

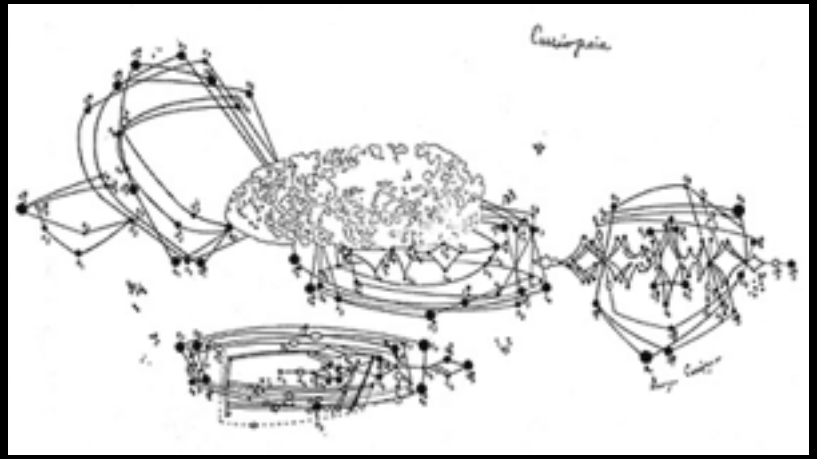
Graphical notation

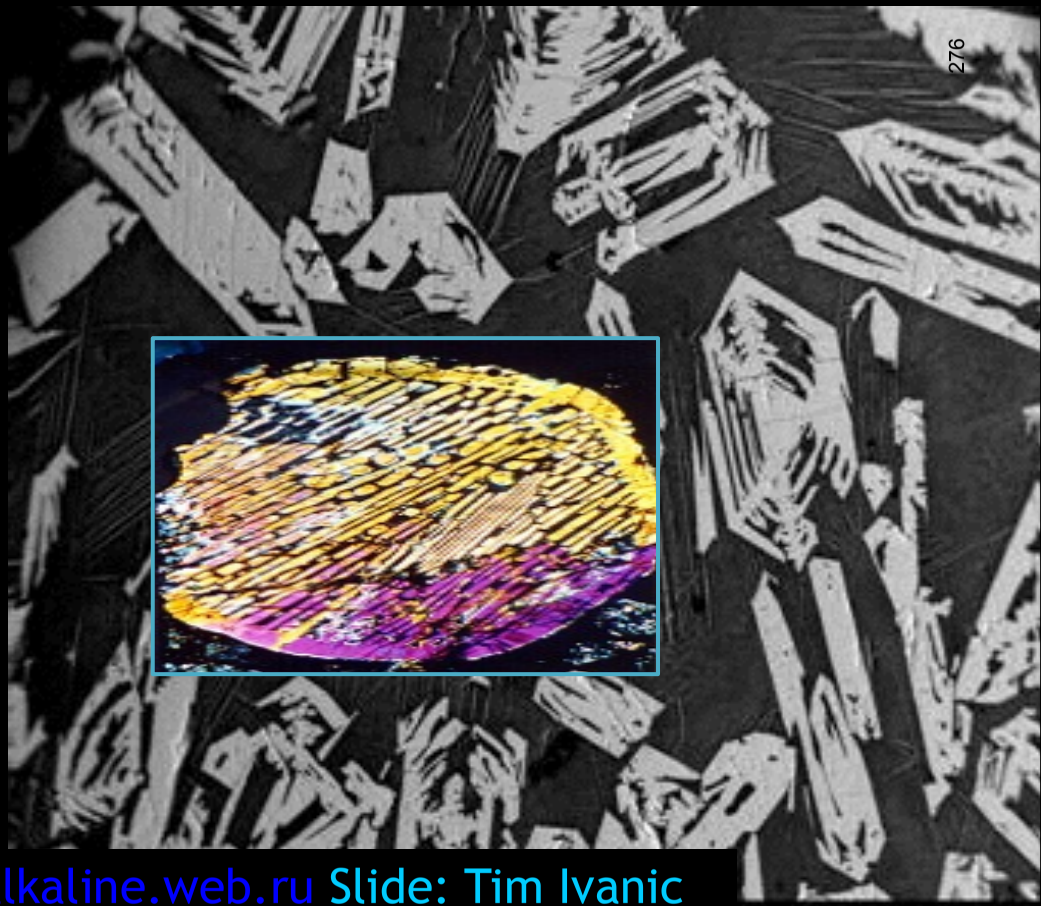




http://dataisnature.com

Geological imagery as graphic notation





http://dataisnature.com http://www.alkaline.web.ru Slide: Tim Ivanic

Eonsounds: Fiamignano Gorge

- Based on a dataset from the Fiamignano area, Apennines, Italy.
- Composers limited themselves to working with a small number of parameters: downstream distribution and stream power as measured at different locations.
- Data was used both to construct melodies by transforming data sets into pitch sets using MAX/MSP and also using the data to control synthesis parameters in real time.
- The piece makes extensive use of FM (frequency modulation) synthesis, employing a Korg Volca FM synthesizer, based on the popular Yamaha DX7 synthesizer.

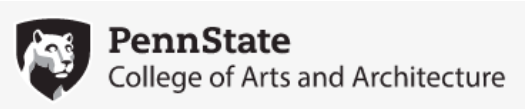
Amatrice earthquake

- 24th Aug 2016, 6.2 M, 297 deaths
- James Cave Composer-in-Residence at the Mahler-LeWitt Studios in Spoleto
- Landscape-reactive performance
- Earth's response, erosion, recovery
- Inspired by Alberto Burri's Cretto di Ghibellina

Earthquake devastation in Amatrice

Cretto di Ghibellina

ICAD 2017 was made possible by support from



The College of Arts and Architecture
(http://artsandarchitecture.psu.edu/)

The Arts & Design Research Incubator
(http://adri.psu.edu/)

The Graduate Program in Acoustics
(http://www.acs.psu.edu/default.aspx)

The College of Information Sciences and Technology
(https://ist.psu.edu/)

The Office of the Vice President for Research
(https://www.research.psu.edu/)

The National Science Foundation
(https://www.nsf.gov/)

The International Community for Auditory Display
(http://icad.org/)